

# Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/JP04/018901

International filing date: 17 December 2004 (17.12.2004)

Document type: Certified copy of priority document

Document details: Country/Office: JP  
Number: 2003-422781  
Filing date: 19 December 2003 (19.12.2003)

Date of receipt at the International Bureau: 20 January 2005 (20.01.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland  
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

PCT/JP 2004/018901

22.12.2004

日本国特許庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 2003年12月19日  
Date of Application:

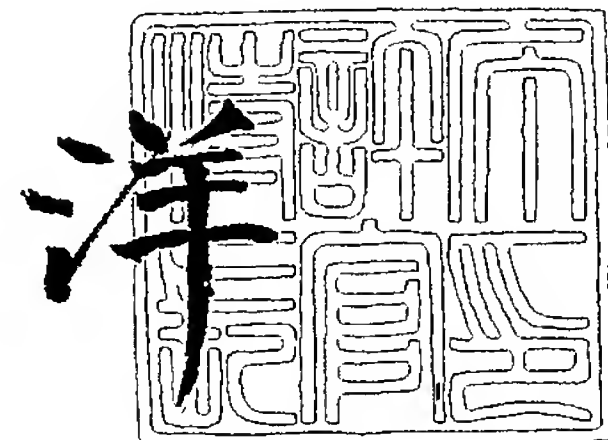
出願番号 特願2003-422781  
Application Number:  
[ST. 10/C]: [JP 2003-422781]

出願人 日本電気株式会社  
Applicant(s):

2004年 9月29日

特許庁長官  
Commissioner,  
Japan Patent Office

小川



出証番号 出証特2004-3087174

【書類名】 特許願  
【整理番号】 34103832  
【提出日】 平成15年12月19日  
【あて先】 特許庁長官 殿  
【国際特許分類】 G01N 33/68  
【発明者】  
    【住所又は居所】 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内  
    【氏名】 鳥居 宏在  
【発明者】  
    【住所又は居所】 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内  
    【氏名】 宮崎 賢司  
【発明者】  
    【住所又は居所】 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内  
    【氏名】 上條 憲一  
【発明者】  
    【住所又は居所】 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内  
    【氏名】 次田 皓  
【特許出願人】  
    【識別番号】 000004237  
    【氏名又は名称】 日本電気株式会社  
【代理人】  
    【識別番号】 100123788  
    【弁理士】  
    【氏名又は名称】 宮崎 昭夫  
    【電話番号】 03-3585-1882  
【選任した代理人】  
    【識別番号】 100088328  
    【弁理士】  
    【氏名又は名称】 金田 暢之  
【選任した代理人】  
    【識別番号】 100106297  
    【弁理士】  
    【氏名又は名称】 伊藤 克博  
【選任した代理人】  
    【識別番号】 100106138  
    【弁理士】  
    【氏名又は名称】 石橋 政幸  
【手数料の表示】  
    【予納台帳番号】 201087  
    【納付金額】 21,000円  
【提出物件の目録】  
    【物件名】 特許請求の範囲 1  
    【物件名】 明細書 1  
    【物件名】 図面 1  
    【物件名】 要約書 1  
    【包括委任状番号】 0304683

## 【書類名】 特許請求の範囲

## 【請求項 1】

解析対象のタンパク質について、  
公知タンパク質に関するデータ・ベースに収録される、公知タンパク質個々に関して、該公知タンパク質を構成するペプチド鎖の完全長アミノ酸配列をコードしているゲノム遺伝子、および該完全長アミノ酸配列の翻訳を可能とする、mRNA中の読み枠の塩基配列、ならびに、その塩基配列でコードされる（推断）完全長アミノ酸配列の配列情報を参照して、前記解析対象タンパク質において実測される質量分析の結果に基づき、該解析対象タンパク質に相当すると推定される、前記データ・ベースに収録される、公知タンパク質の一つを選定する方法であって、

(1) 前記解析対象タンパク質において実測される質量分析の結果は、

予め単離した当該対象タンパク質を構成するペプチド鎖に対して、Cys-Cys結合が存在する際に、該ジスルフィド(S-S)結合を開裂させることが可能な還元処理を施し、且つ、該対象タンパク質の折り畳み(folding)を解き、該対象タンパク質を構成するペプチド鎖を直鎖状にする処理を行い、

さらに、特定のアミノ酸、またはアミノ酸配列において、選択的にペプチド鎖を切断する、部位特異的なタンパク質分解処理を施し、該対象タンパク質より採取される、前記直鎖状ペプチド鎖に由来するペプチド断片複数を等量的かつ選択的に調製し、

少なくとも、生成する前記ペプチド断片複数の質量(M)について、対応する一価の「親の陽イオン種」の分子量( $M+H/Z$ ;  $Z=1$ )、または一価の「親の陰イオン種」の分子量( $M-H/Z$ ;  $Z=1$ )として、質量分析法で測定される結果に基づき、決定されるペプチド断片複数の各質量実測値(Mex)のセットを含む質量分析の結果であり、

(2) 前記公知タンパク質に関するデータ・ベースに収録される、公知タンパク質個々に関して、該公知タンパク質を構成するペプチド鎖の完全長アミノ酸配列をコードしているゲノム遺伝子、および該完全長アミノ酸配列への翻訳を可能とする、mRNA中の読み枠の塩基配列、ならびに、その塩基配列でコードされる（推断）完全長アミノ酸配列の配列情報を参照して、

該公知タンパク質をコードしているゲノム遺伝子に基づき、翻訳される前記完全長アミノ酸配列を有するペプチド鎖に対して、前記Cys側鎖上のスルファニル基(-SH)に対する還元処理と、前記部位特異的なタンパク質分解処理とを施した際に生成すると予測される、前記完全長アミノ酸配列を有するペプチド鎖に由来するペプチド断片複数の推定分子量(Mref)を算出し、各公知タンパク質に由来する、推定ペプチド断片複数が有する推定分子量(Mref)のセットとし、

前記公知タンパク質に関するデータ・ベースに収録される、公知タンパク質個々に関して算定される、該公知タンパク質に由来する、推定ペプチド断片複数が有する推定分子量(Mref)のセット全数から構成される、ペプチド断片複数の推定分子量(Mref)のデータ・セットを、参照基準データ・ベースとして利用し、

(3) 前記解析対象タンパク質において決定されるペプチド断片複数の各質量実測値(Mex)のセットと、前記公知タンパク質に関するデータ・ベースに収録される、公知タンパク質個々に関して算定される、各公知タンパク質に由来する、推定ペプチド断片複数が有する推定分子量(Mref)のセットとを対比させ、

利用している質量分析法自体に起因する測定誤差を考慮して、各質量実測値(Mex)と、各公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量(Mref)とが実質的に一致すると判定される、該解析対象タンパク質由来の実測ペプチド断片数ならびに該公知タンパク質由来の推定ペプチド断片数を、前記参照基準データ・ベースに含まれる、公知タンパク質毎に決定する一次対比操作を行い、

前記一次対比操作において決定される、各公知タンパク質に対して、一致すると判定される、該解析対象タンパク質由来の実測ペプチド断片数ならびに該公知タンパク質由来の推定ペプチド断片数の多い順に、公知タンパク質を選択し、その一致数が最大を示す公知タンパク質を、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパ



ク質の群とし、

(4) 前記一次候補の公知タンパク質の群に含まれる公知タンパク質種類が一つである場合、前記データ・ベース中より選別される該一種類の公知タンパク質を、該解析対象タンパク質に対する、単一の同定の候補と判定する

ことを特徴とする質量分析法を利用するタンパク質の同定方法。

【請求項 2】

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質の配列情報を参照し、

前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) とは、一致するとは判定されていない、該解析対象タンパク質由来の実測ペプチド断片数が零である場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、確度の高い単一の同定の候補と判定することを特徴とする請求項 1 に記載の方法。

【請求項 3】

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質の配列情報を参照し、

前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致すると判定されている、該解析対象タンパク質由来の実測ペプチド断片複数について、対応している該公知タンパク質に由来する推定ペプチド断片が占めるべき位置に配置した際、一致すると判定される前記実測ペプチド断片の一群が、該公知タンパク質の完全長アミノ酸配列中に含まれる、連続したアミノ酸配列を構成している場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、確度の高い単一の同定の候補と判定することを特徴とする請求項 1 に記載の方法。

【請求項 4】

前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、

該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分と連結される、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定の推定ペプチド断片の群に対して、該未同定の推定ペプチド断片中に存在するアミノ酸残基の側鎖への修飾基付加に起因する翻訳後修飾が存在すると仮定した際に予測される、アミノ酸残基の側鎖への修飾基付加に起因する翻訳後修飾を有する推定ペプチド断片に対する推定分子量 (Mref) を算定し、

前記修飾基付加に起因する翻訳後修飾を有する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記修飾基付加に起因する翻訳後修飾を有する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、確度の高い単一の同定の候補と判定することを特徴とする請求項 3 に記載の方法。

【請求項 5】

前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致する

とは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、

該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分と連結される、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定のN末側に存在する推定ペプチド断片の群部分について、翻訳後、そのN末側の一部を切断されるプロセッシングが生じ、成熟型タンパク質へと変換されると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、翻訳後のN末プロセッシングに由来する推定ペプチド断片複数の推定分子量 (Mref) を算出し、

前記翻訳後のN末プロセッシングに由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記翻訳後のN末プロセッシングに由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、確度の高い単一の同定の候補と判定することを特徴とする請求項 3 に記載の方法。

#### 【請求項 6】

前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数の有する推定分子量 (Mref) と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、

該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分と連結される、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定のC末側に存在する推定ペプチド断片の群の部分について、翻訳後、そのC末側の一部を切断されるプロセッシングが生じ、C末切除型タンパク質へと変換されると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、翻訳後のC末切除型プロセッシングに由来する推定ペプチド断片複数の推定分子量 (Mref) を算出し、

前記翻訳後のC末切除型プロセッシングに由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記翻訳後のC末プロセッシングに由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、確度の高い単一の同定の候補と判定することを特徴とする請求項 3 に記載の方法。

#### 【請求項 7】

前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数の有する推定分子量 (Mref) と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、

該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分と連結される、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定の推定ペプチド断片の群部分をコードするゲノム遺伝子部分において、該ゲノム遺伝子部分に含まれるエキソン複数において、推定されているRNAスプライシングと相違するスプライシングが生じていると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成



すると予測される、別のスプライシングに由来する推定ペプチド断片複数の推定分子量 (Mref) を算出し、

前記別のスプライシングに由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記別のスプライシングに由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、確度の高い単一の同定の候補と判定することを特徴とする請求項 3 に記載の方法。

#### 【請求項 8】

前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数の有する推定分子量 (Mref) と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、

該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分と連結される、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定の推定ペプチド断片の群の部分において、そのアミノ酸配列の一部が除去されるプロテイン・スプライシングが生じていると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、プロテイン・スプライシングに由来する推定ペプチド断片複数の推定分子量 (Mref) を算出し、

前記プロテイン・スプライシングに由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記プロテイン・スプライシングに由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、確度の高い単一の同定の候補と判定することを特徴とする請求項 3 に記載の方法。

#### 【請求項 9】

前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数の有する推定分子量 (Mref) と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、

該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分と連結される、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定の推定ペプチド断片の群に対して、それをコードするゲノム遺伝子部分において、該ゲノム遺伝子部分に含まれるエキソン中に、一塩基多型に起因する、翻訳されるアミノ酸置換が一つ生じていると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、一塩基多型のアミノ酸置換に由来する推定ペプチド断片複数の推定分子量 (Mref) を算出し、

前記一塩基多型のアミノ酸置換に由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記一塩基多型のアミノ酸置換に由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実

測ペプチド断片が少なくとも一つ選別される場合、

前記工程（４）において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、確度の高い単一の同定の候補と判定することを特徴とする請求項 3 に記載の方法。

【請求項 1 0】

前記工程（４）において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質の配列情報を参照し、

前記工程（３）の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量（Mref）と、一致すると判定されている、該解析対象タンパク質由来の実測ペプチド断片複数について、対応している該公知タンパク質に由来する推定ペプチド断片が占めるべき位置に配置した際、一致すると判定される前記実測ペプチド断片の一群は、一部の推定ペプチド断片が占めるべき位置を除き、該公知タンパク質の完全長アミノ酸配列中に含まれる、連続したアミノ酸配列を構成している場合、

前記工程（４）において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、確度の高い同定の候補と判定することを特徴とする請求項 1 に記載の方法。

【請求項 1 1】

前記工程（３）の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量（Mref）と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、

該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分中に存在する、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定の推定ペプチド断片の群に対して、該未同定の推定ペプチド断片中に存在するアミノ酸残基の側鎖への修飾基付加に起因する翻訳後修飾が存在すると仮定した際に予測される、アミノ酸残基の側鎖への修飾基付加に起因する翻訳後修飾を有する推定ペプチド断片に対する推定分子量（Mref）を算定し、

前記修飾基付加に起因する翻訳後修飾を有する推定ペプチド断片に対する推定分子量（Mref）と一致する、質量実測値（Mex）を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記修飾基付加に起因する翻訳後修飾を有する推定ペプチド断片に対する推定分子量（Mref）と一致する、質量実測値（Mex）を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、

前記工程（４）において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、確度の高い単一の同定の候補と判定することを特徴とする請求項 1 0 に記載の方法。

【請求項 1 2】

前記工程（３）の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量（Mref）と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、

該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分中に存在する、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定である、内部未同定領域の推定ペプチド断片の群の部分にコードするゲノム遺伝子部分において、該ゲノム遺伝子部分に含まれるエキソン複数において、推定されている RNA スプライシングと相違するスプライシングが生じていると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、別のスプライシングに由来する推定ペプチド断片複数の推定分子量（Mref）を算出し、



前記別のスプライシングに由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記別のスプライシングに由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、確度の高い単一の同定の候補と判定することを特徴とする請求項 10 に記載の方法。

#### 【請求項 13】

前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、

該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分中に存在する、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定である、内部未同定領域の推定ペプチド断片の群部分において、そのアミノ酸配列の一部が除去されるプロテイン・スプライシングが生じていると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、プロテイン・スプライシングに由来する推定ペプチド断片複数の推定分子量 (Mref) を算出し、

前記プロテイン・スプライシングに由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記プロテイン・スプライシングに由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、確度の高い単一の同定の候補と判定することを特徴とする請求項 10 に記載の方法。

#### 【請求項 14】

前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、

該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分中に存在する、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定である、内部未同定領域の推定ペプチド断片の群の各部分に対して、それコードするゲノム遺伝子部分において、該ゲノム遺伝子部分に含まれるエキソン中に、一塩基多型に起因する、翻訳されるアミノ酸置換が一つ生じていると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、一塩基多型のアミノ酸置換に由来する推定ペプチド断片複数の推定分子量 (Mref) を算出し、

前記一塩基多型のアミノ酸置換に由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記一塩基多型のアミノ酸置換に由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候



補と判定される公知タンパク質は、確度の高い単一の同定の候補と判定することを特徴とする請求項 10 に記載の方法。

【請求項 15】

少なくとも、上述する二次対比操作に際して、  
前記解析対象タンパク質において実測される質量分析の結果として、

生成する前記ペプチド断片複数の質量 ( $M$ ) について、対応する一価の「親の陽イオン種」の分子量 ( $M+H/Z$ ;  $Z=1$ )、または一価の「親の陰イオン種」の分子量 ( $M-H/Z$ ;  $Z=1$ ) として、質量分析法で測定される結果に基づき、決定されるペプチド断片複数の各質量実測値 ( $M_{ex}$ ) のセットに加えて、

少なくとも、一次対比操作において、未同定の該解析対象タンパク質由来実測ペプチド断片と判定された該解析対象タンパク質由来の実測ペプチド断片に関して、MS/MS 分析法により、該ペプチド断片の前記「親の陽イオン種」に由来する「娘イオン種」として、フラグメント化された派生イオン種の分子量、または、該ペプチド断片の前記「親の陰イオン種」に由来する「娘イオン種」として、フラグメント化された派生イオン種の分子量の測定結果をも利用し、

該二次対比操作において、前記推定ペプチド断片に対する推定分子量 ( $M_{ref}$ ) と一致する、質量実測値 ( $M_{ex}$ ) を有する未同定の該解析対象タンパク質由来実測ペプチド断片として、新たに選別された解析対象タンパク質由来の実測ペプチド断片に関して、

対応する前記推定ペプチド断片を構成する、想定のアミノ酸配列ならびに付加的な修飾基に起因して、MS/MS 分析において生成すると予測される、フラグメント化された派生イオン種の分子量に関しても、該解析対象タンパク質由来の実測ペプチド断片に対する、前記フラグメント化された派生イオン種の分子量の実測結果との対比を行い、

少なくとも、該解析対象タンパク質由来の実測ペプチド断片に対する、前記フラグメント化された派生イオン種の分子量の実測結果と、対応する前記推定ペプチド断片において予測される、フラグメント化された派生イオン種の分子量の予測値との間においても、対応関係が確認される場合、

前記二次対比操作における、選別される前記解析対象タンパク質由来の実測ペプチド断片は、高い確度の判定とし、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、確度の高い単一の同定の候補と判定することを特徴とする請求項 4～9 および 11～14 のいずれか一項に記載の方法。

【請求項 16】

前記部位特異的なタンパク質分解処理に先立ち、該直鎖状のペプチド鎖に対して、Cys 側鎖上のスルファニル基 ( $-SH$ ) に対する選択的な保護基の導入を行い、得られる Cys 保護処理済みの直鎖状ペプチド鎖とすることを特徴とする請求項 1～15 のいずれか一項に記載の方法。

となっている。加えて、タンパク質自体のペプチド鎖C末端の部分的アミノ酸配列に関しても、例えば、国際公開 WO 03/081255 A1号パンフレットに開示される「ペプチドのC末端アミノ酸配列解析方法」の手法などを利用して、質量分析法によって、高い確度で解析が可能となっている。従って、仮に、公知のタンパク質候補の標準試料が入手可能であれば、対比させる一群のペプチド断片の各分子量に関する実測データも利用可能であり、前記の質量分析法で得られる情報に基づき、相当の信頼度で、解析対象のタンパク質と公知のタンパク質候補とが同一であるか否かを判定することが可能である。

#### 【0010】

現実的には、既に報告されている公知タンパク質に関するデータ・ベースに収録される各タンパク質に関して、標準試料を利用可能なものは少なく、開示されているアミノ酸配列の情報に関しても、対応する遺伝子情報、すなわち、これらのペプチドをコードしているゲノム遺伝子やm-RNAより調製されたc-DNAの塩基配列に基づき、翻訳されるペプチド鎖の推断されるアミノ酸配列情報がその相当部分を占めている。例えば、成熟型タンパク質へと変換される翻訳後の「プロセッシング」を受けるタンパク質に関して、実際に、シグナル・ペプチダーゼによって切断され、シグナル・ペプチド部の部分アミノ酸配列に関して、その詳細は不明であるような、部分的に不完全さを残す、アミノ酸配列情報しか開示されていないものも多く含まれている。

#### 【0011】

そのため、公知のタンパク質候補に対して、部位特異的なタンパク質分解酵素を作用させ、生成する一群のペプチド断片の各分子量に関する実測データに代えて、対応する遺伝子情報、すなわち、その完全長のペプチド鎖をコードしているゲノム遺伝子やm-RNAより調製されたc-DNAの塩基配列に基づき、翻訳されるペプチド鎖の推断されるアミノ酸配列情報に基づき、部位特異的なタンパク質分解酵素消化で生成すると予測される、一群のペプチド断片のアミノ酸配列部分に対応する各式量（推定分子量）を参照基準として、解析対象のタンパク質の質量分析で実測される、一群のペプチド断片の各分子量と対比させ、高い一致を示すか否かに基づき、解析対象のタンパク質と公知のタンパク質候補とが同一であるか否かを、相当の信頼度で判定する手法の開発、研究が、現在精力的に進められている。

#### 【0012】

特には、解析対象のタンパク質が、上述する「翻訳後修飾」を有するタンパク質、スプライシング変異体型タンパク質あるいはプロテインースプライシング変異体型タンパク質であった場合、それらをコードしているゲノム遺伝子やm-RNAより調製されたc-DNAの塩基配列に基づき、翻訳されるペプチド鎖の推断されるアミノ酸配列情報は、理想的な完全長のペプチド鎖に対するものであり、参照基準として利用する、部位特異的なタンパク質分解酵素消化で生成すると予測される、一群のペプチド断片のアミノ酸配列部分に対応する各式量（推定分子量）と、解析対象のタンパク質の質量分析で実測される、一群のペプチド断片の各分子量と対比させた際、一致の見られないペプチド断片が存在することになる。また、解析対象のタンパク質が、「一塩基多型」に由来して、翻訳されるペプチド鎖のアミノ酸配列に変異が生じている、所謂、「表現型」の異なる変異体タンパク質であった場合にも、報告されている公知のタンパク質候補の有する「標準」のアミノ酸配列情報から予測される、一群のペプチド断片のアミノ酸配列部分に対応する各式量（推定分子量）と、解析対象のタンパク質の質量分析で実測される、一群のペプチド断片の各分子量と対比させた際、やはり、一致の見られないペプチド断片が存在することになる。

#### 【0013】

換言するならば、報告されている公知のタンパク質候補の有する「標準」のアミノ酸配列情報から予測される、一群のペプチド断片のアミノ酸配列部分に対応する各式量（推定分子量）と、解析対象のタンパク質の質量分析で実測される、一群のペプチド断片の各分子量と対比させた際、相当数のペプチド断片に関しては、実測される分子量（Mex）と参照基準の推定分子量（Mref）とが一致している際、一致の見られないペプチド断片について、その不一致を引き起こす要因を合理的に推定することが可能となれば、解析対象の



タンパク質の同定、すなわち、それをコードしている遺伝子から翻訳されるべき、公知のタンパク質候補の特定、ならびに、その不一致を引き起こす要因の高い蓋然性での推断とが可能となる。すなわち、解析対象のタンパク質が、例えば、「翻訳後修飾」を有するタンパク質、あるいは、ある公知タンパク質に対する、スプライシング変異体、「一塩基多型」変異体などに相当する場合、一致の見られないペプチド断片をもたらす要因である、かかる「翻訳後修飾」やアミノ酸配列上の変異を解析する際、基準とすべき「公知タンパク質の候補」を高い蓋然性で特定することが可能となる。

#### 【0014】

なお、本発明においては、実際に、その存在が確認され、報告されている、「狭義の公知タンパク質」に加えて、現実には、タンパク質自体の存在は確認されていないものの、その翻訳に利用される mRNA の存在が確認され、報告されている、「発現が公知となっているタンパク質」、ならびに、mRNA の存在は確認されていないものの、ゲノム遺伝子解析の結果、前駆体 RNA への転写、その後、mRNA から完全長のペプチド鎖への翻訳が可能なコード遺伝子と推定され、データ・ベース上にかかるコード遺伝子が収録されている、「コード遺伝子が公知となっているタンパク質」のように、コード遺伝子の塩基配列、ならびに翻訳されるペプチド鎖の推断されるアミノ酸配列情報が、データ・ベース上に収録され、公知となっている「広義の意味で公知となっているタンパク質」を含めてある、スプライシング変異体型タンパク質が実際に確認され、データ・ベース上にかかるコード遺伝子が収録されている、「コード遺伝子が公知となっているタンパク質」のように、ゲノム上のコード遺伝子の塩基配列、ならびに mRNA から翻訳されるペプチド鎖の推断されるアミノ酸配列情報が報告されている場合には、かかる公知のスプライシング変異体型タンパク質も、「公知タンパク質」に含まれる。

#### 【0015】

本発明は、前記の課題を解決するもので、本発明の目的は、解析対象のタンパク質について、一旦、該タンパク質を単離した上で、特定のアミノ酸、またはアミノ酸配列において、選択的にペプチド鎖を切断する、部位特異的なタンパク質分解処理を施して生成する、該タンパク質由来の一群のペプチド断片について、質量分析法により実測される一群のペプチド断片の各分子量測定結果に基づき、利用可能な公知のタンパク質に関する、それらをコードしているゲノム遺伝子や mRNA より調製された cDNA の塩基配列、ならびに、該コード塩基配列に従って、翻訳される完全長のペプチド鎖の推断されるアミノ酸配列情報のデータ・ベースを参照し、その推定されるアミノ酸配列を有する完全長のペプチド鎖に対して、前記部位特異的なタンパク質分解処理を施すことで生成すると予測される、一群のペプチド断片のアミノ酸配列部分に対応する各式量（推定分子量）を参照基準として利用し、実測される分子量（Mex）と参照基準の推定分子量（Mref）とが一致しているペプチド断片数の多寡を一次判断基準として、それをコードしている遺伝子から翻訳されるべき、公知のタンパク質候補の特定、ならびに、該特定される公知のタンパク質候補を、その遺伝子産物とする公知の遺伝子候補の特定と、仮に、一致の見られないペプチド断片が見出される際には、その不一致を引き起こす要因を高い蓋然性での推断することを可能とする、質量分析法を利用して、タンパク質の同定を行う新規な解析手法を提することにある。より具体的には、本発明の目的は、前記一次判断基準に基づき、公知のゲノム遺伝子や mRNA より調製された cDNA の塩基配列、ならびに、該コード塩基配列に従って、翻訳される完全長のペプチド鎖の推断されるアミノ酸配列情報のデータ・ベースから選別される公知のタンパク質候補に対して、解析対象のタンパク質が、翻訳後修飾を有するタンパク質、スプライシング変異体型タンパク質、あるいは、一塩基多型に由来する表現型の異なる変異体タンパク質に相当する際、実際に見出される、実測される分子量（Mex）と参照基準の推定分子量（Mref）との間に一致の見られないペプチド断片に関して、その不一致を引き起こす要因が、それぞれ、翻訳後修飾を有するタンパク質、スプライシング変異体型タンパク質、あるいは、一塩基多型に由来する表現型の異なる変異体タンパク質であることを、高い蓋然性での推断することを可能とする解析手法

を提供することにある。

【課題を解決するための手段】

【0016】

本発明者らは、上記の課題を解決すべく、鋭意研究を進めたところ、例えば、同定を行う対象のタンパク質に関して、一旦、電気泳動法などの分離手段を利用して、原試料中から単離した上で、さらに、該対象タンパク質の折り畳み (folding) を解き、同時に、必要に応じて、構成するペプチド鎖の鎖間ならびに鎖内の Cys-Cys 結合を還元処理して、そのジスルフィド (S-S) 結合を開裂させ、該対象タンパク質を構成するペプチド鎖を直鎖状にして、構成する直鎖状ペプチド鎖複数をそれぞれ、分離・回収する。

次いで、各直鎖状ペプチド鎖に対して、特定のアミノ酸、またはアミノ酸配列において、選択的にペプチド鎖を切断する、部位特異的なタンパク質分解処理を施すことで、対象となるタンパク質を構成するペプチド鎖に由来するペプチド断片を選択的に作成でき、ペプチドの分析に適する MALDI-TOF-MS などの質量分析法を利用することで、生成するペプチド断片複数の質量を、対応する一価の「親の陽イオン種」の分子量 ( $M+H/Z$ ;  $Z=1$ )、ならびに一価の「親の陰イオン種」の分子量 ( $M-H/Z$ ;  $Z=1$ ) として、高い精度で測定した結果に基づき、観測されるペプチド断片複数の質量実測値 (Mex) を決定することが可能となることを確認した。

【0017】

一方、既に報告されている公知タンパク質に関するデータ・ベースに収録される各タンパク質に関して、例えば、各タンパク質を構成するペプチド鎖の完全長アミノ酸配列をコードしているゲノム遺伝子、および該完全長アミノ酸配列の翻訳を可能とする、mRNA 中の読み枠の塩基配列、ならびに、その塩基配列でコードされる (推断) 完全長アミノ酸配列の配列情報に基づき、

該公知タンパク質を構成するペプチド鎖について、前記完全長アミノ酸配列を有する、その翻訳時におけるペプチド鎖に対して、

前記の直鎖化処理、ならびに部位特異的なタンパク質分解処理を施した際、すなわち、完全長アミノ酸配列を有するペプチド鎖に含まれる、予め、Cys-Cys 結合を Cys 側鎖上のスルファニル基 (-SH) へと還元する処理を施し、直鎖状のペプチド鎖への変換を行った上で、特定のアミノ酸、またはアミノ酸配列において、選択的にペプチド鎖を切断する、部位特異的なタンパク質分解処理を施した際に生成すると予測される、ペプチド断片複数の推定分子量 (Mref) を算出することができる。

【0018】

このデータ・ベースに収録される各公知タンパク質に関する配列情報に基づき、算出される、各公知タンパク質より生成されると予測される、参照基準とするペプチド断片複数の推定分子量 (Mref) のデータ・セットと、上記対象タンパク質において決定される、ペプチド断片複数の実測分子量 (Mex) のデータ・セットとを、対比させることにより、

利用している質量分析法自体に起因する測定誤差を考慮して、実質的に一致すると判定されるペプチド断片数が、参照基準とする、公知タンパク質毎に決定される。この一次対比操作において、各公知タンパク質に対して、「一致」したと判定される「実測」ペプチド断片数、ならびに、「一致」と判定されていない「実測」ペプチド断片の選別がなされ、「一致」したと判定される「実測」ペプチド断片数が多い順に、公知タンパク質を選択して、該解析対象のタンパク質に対する、その同定の候補とする「一次候補の公知タンパク質」群とすることができる。

【0019】

仮に、この一次対比操作段階において、

(A) 「一致」と判定されていない「実測」ペプチド断片が残っていない場合、あるいは、選択された「一次候補の公知タンパク質」の完全長アミノ酸配列を参照して、「一致」



したと判定される「実測」ペプチド断片を、対応している、「一次候補の公知タンパク質」由来の「推定」ペプチド断片が占めるべき位置に配置した際、「一致」したと判定される「実測」ペプチド断片の一群が、連続したアミノ酸配列を構成していると判断される場合、解析対象のタンパク質は、選択された「一次候補の公知タンパク質」に相当すると、高い確度で同定が可能であることが判明した。

#### 【0020】

また、「一致」と判定されていない「実測」ペプチド断片が残っている場合であっても、(B-1) 選択された「一次候補の公知タンパク質」の完全長アミノ酸配列を参照して、「一致」したと判定される「実測」ペプチド断片を、対応している、「一次候補の公知タンパク質」由来の「推定」ペプチド断片が占めるべき位置に配置した際、「一致」したと判定される「実測」ペプチド断片の一群が、連続したアミノ酸配列を構成していると判断される場合、解析対象のタンパク質は、選択された「一次候補の公知タンパク質」に相当する、ないしは、該選択された「一次候補の公知タンパク質」をコードしている遺伝子の産物であると、高い確度で同定が可能であることが判明した。

#### 【0021】

この(B-1)の場合、「一致」と判定されていない「実測」ペプチド断片に関して、一次同定された「一次候補の公知タンパク質」に由来し、前記判断において、特定されている「連続したアミノ酸配列」部分に連結される、未同定の「推定」ペプチド断片群より、

(B-1-1) 翻訳後修飾に起因して、未同定の「推定」ペプチド断片が有する推定分子量(Mref)と相違する、質量実測値(Mex)を有する「実測」ペプチド断片が生成していること；

(B-1-2) 「一次候補の公知タンパク質」において、予測されているスプライシング過程とは、相違するスプライシングが発生していることに起因して、未同定の「推定」ペプチド断片が有する推定分子量(Mref)と相違する、質量実測値(Mex)を有する「実測」ペプチド断片が生成していること；

(B-1-3) 「一次候補の公知タンパク質」における(推断)完全長アミノ酸配列と、未同定の「推定」ペプチド断片群において、「一塩基多型」に付随して、アミノ酸置換が発生していることに起因して、未同定の「推定」ペプチド断片が有する推定分子量(Mref)と相違する、質量実測値(Mex)を有する「実測」ペプチド断片が生成していること；

のいずれかが原因となって、「一致」と判定されていない「実測」ペプチド断片が残っていると推断されると、解析対象のタンパク質は、選択された「一次候補の公知タンパク質」に相当する、ないしは、該選択された「一次候補の公知タンパク質」をコードしている遺伝子の産物であると、より高い確度で同定が可能であることが判明した。

#### 【0022】

加えて、「一致」と判定されていない「実測」ペプチド断片が残っている場合、(B-2) 選択された「一次候補の公知タンパク質」の完全長アミノ酸配列を参照して、「一致」したと判定される「実測」ペプチド断片を、対応している、「一次候補の公知タンパク質」由来の「推定」ペプチド断片が占めるべき位置に配置した際、「一致」したと判定される「実測」ペプチド断片の一群は、一部の「推定」ペプチド断片が占めるべき位置を除き、連続したアミノ酸配列を構成していると判断される場合、解析対象のタンパク質は、選択された「一次候補の公知タンパク質」に相当すると、ある程度高い確度で同定が可能であることが判明した。

#### 【0023】

この(B-2)の場合、「一致」と判定されていない「実測」ペプチド断片に関して、一次同定された「一次候補の公知タンパク質」に由来し、前記判断において、特定されている「連続したアミノ酸配列」の内部において、「一致」と判定される「実測」ペプチド断片が特定されず、内部の未同定領域に相当する「推定」ペプチド断片の群に関して、

(B-2-1) 翻訳後修飾に起因して、内部未同定領域の「推定」ペプチド断片が有する



推定分子量 (Mref) と相違する、質量実測値 (Mex) を有する「実測」ペプチド断片が生成していること；

(B-2-2) 「一次候補の公知タンパク質」において、予測されているスプライシング過程とは、相違するスプライシングが発生していることに起因して、内部未同定領域の「推定」ペプチド断片が有する推定分子量 (Mref) と相違する、質量実測値 (Mex) を有する「実測」ペプチド断片が生成していること；

(B-2-3) 「一次候補の公知タンパク質」における (推断) 完全長アミノ酸配列と、未同定の「推定」ペプチド断片群において、「一塩基多型」に付随して、アミノ酸置換が発生していることに起因して、内部未同定領域の「推定」ペプチド断片が有する推定分子量 (Mref) と相違する、質量実測値 (Mex) を有する「実測」ペプチド断片が生成していること；

のいずれかが原因となって、「一致」と判定されていない「実測」ペプチド断片が残っていると推断されると、解析対象のタンパク質は、選択された「一次候補の公知タンパク質」に相当する、ないしは、該選択された「一次候補の公知タンパク質」をコードしている遺伝子に由来する産物であると、より高い確度で同定が可能であることが判明した。本発明者らは、上記の一連の知見に基づき、本発明を完成するに至った。

#### 【0024】

すなわち、本発明にかかる質量分析法を利用するタンパク質の同定方法は、

解析対象のタンパク質について、

公知タンパク質に関するデータ・ベースに収録される、公知タンパク質個々に関して、該公知タンパク質を構成するペプチド鎖の完全長アミノ酸配列をコードしているゲノム遺伝子、および該完全長アミノ酸配列の翻訳を可能とする、mRNA中の読み枠の塩基配列、ならびに、その塩基配列でコードされる (推断) 完全長アミノ酸配列の配列情報を参照して、前記解析対象タンパク質において実測される質量分析の結果に基づき、該解析対象タンパク質に相当すると推定される、前記データ・ベースに収録される、公知タンパク質の一つを選定する方法であって、

(1) 前記解析対象タンパク質において実測される質量分析の結果は、

予め単離した当該対象タンパク質を構成するペプチド鎖に対して、Cys-Cys結合が存在する際に、該ジスルフィド (S-S) 結合を開裂させることが可能な還元処理を施し、且つ、該対象タンパク質の折り畳み (folding) を解き、該対象タンパク質を構成するペプチド鎖を直鎖状にする処理を行い、

さらに、特定のアミノ酸、またはアミノ酸配列において、選択的にペプチド鎖を切断する、部位特異的なタンパク質分解処理を施し、該対象タンパク質より採取される、前記直鎖状ペプチド鎖に由来するペプチド断片複数を等量的かつ選択的に調製し、

少なくとも、生成する前記ペプチド断片複数の質量 (M) について、対応する一価の「親の陽イオン種」の分子量 ( $M+H/Z$ ;  $Z=1$ )、または一価の「親の陰イオン種」の分子量 ( $M-H/Z$ ;  $Z=1$ ) として、質量分析法で測定される結果に基づき、決定されるペプチド断片複数の各質量実測値 (Mex) のセットを含む質量分析の結果であり、

(2) 前記公知タンパク質に関するデータ・ベースに収録される、公知タンパク質個々に関して、該公知タンパク質を構成するペプチド鎖の完全長アミノ酸配列をコードしているゲノム遺伝子、および該完全長アミノ酸配列への翻訳を可能とする、mRNA中の読み枠の塩基配列、ならびに、その塩基配列でコードされる (推断) 完全長アミノ酸配列の配列情報を参照して、

該公知タンパク質をコードしているゲノム遺伝子に基づき、翻訳される前記完全長アミノ酸配列を有するペプチド鎖に対して、前記Cys側鎖上のスルファニル基 (-SH) に対する還元処理と、前記部位特異的なタンパク質分解処理とを施した際に生成すると予測される、前記完全長アミノ酸配列を有するペプチド鎖に由来するペプチド断片複数の推定分子量 (Mref) を算出し、各公知タンパク質に由来する、推定ペプチド断片複数が有する推定分子量 (Mref) のセットとし、

前記公知タンパク質に関するデータ・ベースに収録される、公知タンパク質個々に関し

て算定される、該公知タンパク質に由来する、推定ペプチド断片複数が有する推定分子量 (Mref) のセット全数から構成される、ペプチド断片複数の推定分子量 (Mref) のデータ・セットを、参照基準データ・ベースとして利用し、

(3) 前記解析対象タンパク質において決定されるペプチド断片複数の各質量実測値 (Mex) のセットと、前記公知タンパク質に関するデータ・ベースに収録される、公知タンパク質個々に関して算定される、各公知タンパク質に由来する、推定ペプチド断片複数が有する推定分子量 (Mref) のセットとを対比させ、

利用している質量分析法自体に起因する測定誤差を考慮して、各質量実測値 (Mex) と、各公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) とが実質的に一致すると判定される、該解析対象タンパク質由来の実測ペプチド断片数ならびに該公知タンパク質由来の推定ペプチド断片数を、前記参照基準データ・ベースに含まれる、公知タンパク質毎に決定する一次対比操作を行い、

前記一次対比操作において決定される、各公知タンパク質に対して、一致すると判定される、該解析対象タンパク質由来の実測ペプチド断片数ならびに該公知タンパク質由来の推定ペプチド断片数の多い順に、公知タンパク質を選択し、その一致数が最大を示す公知タンパク質を、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質の群とし、

(4) 前記一次候補の公知タンパク質の群に含まれる公知タンパク質種類が一つである場合、前記データ・ベース中より選別される該一種類の公知タンパク質を、該解析対象タンパク質に対する、単一の同定の候補と判定することを特徴とする質量分析法を利用するタンパク質の同定方法である。

#### 【0025】

その際、前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質の配列情報を参照し、

前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) とは、一致するとは判定されていない、該解析対象タンパク質由来の実測ペプチド断片数が零である場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0026】

あるいは、前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質の配列情報を参照し、

前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致すると判定されている、該解析対象タンパク質由来の実測ペプチド断片複数について、対応している該公知タンパク質に由来する推定ペプチド断片が占めるべき位置に配置した際、一と判定されると判定される前記実測ペプチド断片の一群が、該公知タンパク質の完全長アミノ酸配列中に含まれる、連続したアミノ酸配列を構成している場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0027】

加えて、前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分と連結される、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチ



ド断片が未同定の推定ペプチド断片の群に対して、該未同定の推定ペプチド断片中に存在するアミノ酸残基の側鎖への修飾基付加に起因する翻訳後修飾が存在すると仮定した際に予測される、アミノ酸残基の側鎖への修飾基付加に起因する翻訳後修飾を有する推定ペプチド断片に対する推定分子量 (Mref) を算定し、

前記修飾基付加に起因する翻訳後修飾を有する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記修飾基付加に起因する翻訳後修飾を有する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0028】

あるいは、前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分と連結される、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定のN末側に存在する推定ペプチド断片の群の部分について、翻訳後、そのN末側の一部を切断されるプロセッシングが生じ、成熟型タンパク質へと変換されると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、翻訳後のN末プロセッシングに由来する推定ペプチド断片複数の推定分子量 (Mref) を算出し、

前記翻訳後のN末プロセッシングに由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記翻訳後のN末プロセッシングに由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0029】

同じく、前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、

該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分と連結される、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定のC末側に存在する推定ペプチド断片の群の部分について、翻訳後、そのC末側の一部を切断されるプロセッシングが生じ、C末切除型タンパク質へと変換されると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、翻訳後のC末切除型プロセッシングに由来する推定ペプチド断片複数の推定分子量 (Mref) を算出し、

前記翻訳後のC末切除型プロセッシングに由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記翻訳後のC末プロセッシングに由来する推定ペプチド断片に対する推定分子量 (M

ref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0030】

また、前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、

該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分と連結される、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定の推定ペプチド断片の群の部分をコードするゲノム遺伝子部分において、該ゲノム遺伝子部分に含まれるエキソン複数において、推定されている前駆体 RNA スプライシングと相違するスプライシングが生じていると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、別のスプライシングに由来する推定ペプチド断片複数の推定分子量 (Mref) を算出し、

前記別のスプライシングに由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記別のスプライシングに由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0031】

あるいは、前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、

該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分と連結される、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定の推定ペプチド断片の群部分において、そのアミノ酸配列の一部が除去されるプロテイン・スプライシングが生じていると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、プロテイン・スプライシングに由来する推定ペプチド断片複数の推定分子量 (Mref) を算出し、

前記プロテイン・スプライシングに由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記プロテイン・スプライシングに由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0032】

加えて、前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タン



パク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、

該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分と連結される、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定の推定ペプチド断片の群に対して、それをコードするゲノム遺伝子部分において、該ゲノム遺伝子部分に含まれるエキソン中に、一塩基多型に起因する、翻訳されるアミノ酸置換が一つ生じていると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、一塩基多型のアミノ酸置換に由来する推定ペプチド断片複数の推定分子量 (Mref) を算出し、

前記一塩基多型のアミノ酸置換に由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記一塩基多型のアミノ酸置換に由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0033】

一方、前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質の配列情報を参照し、

前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致すると判定されている、該解析対象タンパク質由来の実測ペプチド断片複数について、対応している該公知タンパク質に由来する推定ペプチド断片が占めるべき位置に配置した際、一致すると判定される前記実測ペプチド断片の一群は、一部の推定ペプチド断片が占めるべき位置を除き、該公知タンパク質の完全長アミノ酸配列中に含まれる、連続したアミノ酸配列を構成している場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、高い確度で単一の同定の候補と判定することが可能である。

#### 【0034】

その際、前記工程 (3) の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、

該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分中に存在する、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定の推定ペプチド断片の群に対して、該未同定の推定ペプチド断片中に存在するアミノ酸残基の側鎖への修飾基付加に起因する翻訳後修飾が存在すると仮定した際に予測される、アミノ酸残基の側鎖への修飾基付加に起因する翻訳後修飾を有する推定ペプチド断片に対する推定分子量 (Mref) を算定し、

前記修飾基付加に起因する翻訳後修飾を有する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記修飾基付加に起因する翻訳後修飾を有する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、



前記工程（４）において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

【0035】

また、前記工程（３）の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量（Mref）と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、

該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分中に存在する、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定である、内部未同定領域の推定ペプチド断片の群部分をコードするゲノム遺伝子部分において、該ゲノム遺伝子部分に含まれるエキソン複数において、推定されているRNAスプライシングと相違するスプライシングが生じていると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、別のスプライシングに由来する推定ペプチド断片複数の推定分子量（Mref）を算出し、

前記別のスプライシングに由来する推定ペプチド断片に対する推定分子量（Mref）と一致する、質量実測値（Mex）を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記別のスプライシングに由来する推定ペプチド断片に対する推定分子量（Mref）と一致する、質量実測値（Mex）を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、

前記工程（４）において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

【0036】

あるいは、前記工程（３）の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量（Mref）と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、

該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分中に存在する、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定である、内部未同定領域の推定ペプチド断片の群部分において、そのアミノ酸配列の一部が除去されるプロテイン・スプライシングが生じていると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、プロテイン・スプライシングに由来する推定ペプチド断片複数の推定分子量（Mref）を算出し、

前記プロテイン・スプライシングに由来する推定ペプチド断片に対する推定分子量（Mref）と一致する、質量実測値（Mex）を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記プロテイン・スプライシングに由来する推定ペプチド断片に対する推定分子量（Mref）と一致する、質量実測値（Mex）を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、

前記工程（４）において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

【0037】

加えて、前記工程（３）の一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量（Mref）と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が

残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、

該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分中に存在する、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定である、内部未同定領域の推定ペプチド断片の群の部分に対して、それをコードするゲノム遺伝子部分において、該ゲノム遺伝子部分に含まれるエキソン中に、一塩基多型に起因する、翻訳されるアミノ酸置換が一つ生じていると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、一塩基多型のアミノ酸置換に由来する推定ペプチド断片複数の推定分子量 (Mref) を算出し、

前記一塩基多型のアミノ酸置換に由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記一塩基多型のアミノ酸置換に由来する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0038】

少なくとも、上述する二次対比操作に際して、  
前記解析対象タンパク質において実測される質量分析の結果として、

生成する前記ペプチド断片複数の質量 (M) について、対応する一価の「親の陽イオン種」の分子量 ( $M+H/Z$ ;  $Z=1$ )、または一価の「親の陰イオン種」の分子量 ( $M-H/Z$ ;  $Z=1$ ) として、質量分析法で測定される結果に基づき、決定されるペプチド断片複数の各質量実測値 (Mex) のセットに加えて、

少なくとも、一次対比操作において、未同定の該解析対象タンパク質由来実測ペプチド断片と判定された該解析対象タンパク質由来の実測ペプチド断片に関して、MS/MS 分析法により、該ペプチド断片の前記「親の陽イオン種」に由来する「娘イオン種」として、フラグメント化された派生イオン種の分子量、または、該ペプチド断片の前記「親の陰イオン種」に由来する「娘イオン種」として、フラグメント化された派生イオン種の分子量の測定結果をも利用し、

該二次対比操作において、前記推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片として、新たに選別された解析対象タンパク質由来の実測ペプチド断片に関して、

対応する前記推定ペプチド断片を構成する、想定のアミノ酸配列ならびに付加的な修飾基に起因して、MS/MS 分析において生成すると予測される、フラグメント化された派生イオン種の分子量に関しても、該解析対象タンパク質由来の実測ペプチド断片に対する、前記フラグメント化された派生イオン種の分子量の実測結果との対比を行い、

少なくとも、該解析対象タンパク質由来の実測ペプチド断片に対する、前記フラグメント化された派生イオン種の分子量の実測結果と、対応する前記推定ペプチド断片において予測される、フラグメント化された派生イオン種の分子量の予測値との間においても、対応関係が確認される場合、

前記二次対比操作における、選別される前記解析対象タンパク質由来の実測ペプチド断片は、高い確度の判定とし、

前記工程 (4) において、選別される該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0039】

なお、本発明の方法では、前記部位特異的なタンパク質分解処理に先立ち、該直鎖状のペプチド鎖に対して、Cys 側鎖上のスルファニル基 ( $-SH$ ) に対する選択的な保護基



の導入を行い、得られる C y s 保護処理済みの直鎖状ペプチド鎖とすることもでき、その際、推定ペプチド断片においても、C y s 側鎖上のスルファニル基 (—S H) に対する選択的な保護基の導入がなされていると仮定して、その推定分子量を算定する。

【発明の効果】

【0 0 4 0】

本発明にかかる質量分析法を利用するタンパク質の同定方法は、特に、後述する解析対象タンパク質を構成するペプチド鎖が、データ・ベースに収録されている、対応するゲノム遺伝子上にコードされる完全長アミノ酸配列を有するペプチド鎖と対比して、種々の要因に起因する特異的な質量変化を示す場合においても、解析対象タンパク質に対して、質量分析法で得られる情報に基づき、公知タンパク質に関するデータ・ベースに収録される、公知タンパク質個々に関して、該公知タンパク質を構成するペプチド鎖の完全長アミノ酸配列をコードしているゲノム遺伝子、および該完全長アミノ酸配列への翻訳を可能とする、m R N A 中の読み枠の塩基配列、ならびに、その塩基配列でコードされる（推断）完全長アミノ酸配列の配列情報を参照して、該解析対象タンパク質に相当すると推定される、前記データ・ベースに収録される、公知タンパク質の一つを高い確度で選定する方法となる。換言するならば、解析対象のタンパク質が、例えば、「翻訳後修飾」を有するタンパク質、あるいは、ある公知タンパク質に対する、スプライシング変異体、「一塩基多型」変異体などに相当する場合、一致の見られないペプチド断片をもたらす要因である、かかる「翻訳後修飾」やアミノ酸配列上の変異を解析する際、基準とすべき「公知タンパク質の候補」を高い蓋然性で特定することが可能な手段となる。

【発明を実施するための最良の形態】

【0 0 4 1】

生体試料中に含有されるタンパク質が、真核生物、特に、ヒトを代表とする哺乳動物に由来する内因性タンパク質である場合、そのゲノム遺伝子から転写される前駆体 R N A 鎖から、その中に含まれるイントロン部分を前駆体 R N A スプライシング過程で除去し、複数のエキソン領域が、その読み枠を一致させて連結されたコード塩基配列を有する m R N A が作製される。この m R N A から翻訳されるペプチド鎖は、前記コード塩基配列によってコードされている、所謂、完全長のアミノ酸配列を有する形態である。

【0 0 4 2】

公知タンパク質のアミノ酸配列に関して、実際に、それを構成するペプチド鎖のアミノ酸を完全に分析することで、全アミノ酸配列の解明がなされたものは、少数であり、多くのものは、かかる公知タンパク質の生合成において、ペプチド鎖の翻訳に利用される m R N A、または、該 m R N A を鋳型として作製される c D N A の塩基配列解析、あるいは、前記 m R N A の作製において、元となる前駆体 R N A 鎖への転写がなされるゲノム遺伝子の塩基配列解析を利用して、開始コドンから終止コドンに至る、一連のアミノ酸配列へと翻訳可能な読み枠を解明することで、（推断）完全長アミノ酸配列として、特定されている。特に、昨今、ゲノム解析結果に基づき、生体内において、翻訳されると推定される、

（推断）完全長アミノ酸配列と、それをコードするゲノム遺伝子の塩基配列、その翻訳領域を構成する一連のエキソンの一群と、そのエキソン間を区分するイントロン領域に関する塩基配列の情報とを、一体化したデータ・ベースの利用が可能となっている。

【0 0 4 3】

同時に、生化学的研究、あるいは、病理学的な研究成果に基づき、実際に、生体内に存在する形状、例えば、翻訳後、完全長アミノ酸配列を有するペプチド鎖の N 末端に存在するシグナル・ペプチド部などがプロセッシング過程で切除され、成熟型タンパク質となるもの、あるいは、種々の核内移行型タンパク質、例えば、転写因子タンパク質などにおいては、核内移行段階で、特定のアミノ酸残基に対するリン酸化が生じ、また、その後、脱リン酸化、あるいは、核膜透過過程において、更なるプロセッシングを受けた形態となるなど、翻訳後に生じる修飾も少なからず解明されてきている。但し、それらの翻訳後に生じる修飾に関する情報は、前記配列情報に関するデータ・ベースに、付加的な情報として、収録はなされているものではない。

## 【0044】

その他、頻度は低いものの、前駆体RNA鎖から、その中に含まれるイントロン部分を除去してmRNAを作製する、前駆体RNAースプライシング過程において、スプライシングが生じる部位が複数存在し、その複数の選択肢から、個体、状況などの決定因子に依存して、異なる種類のスプライシングが選択的に起こる、alternative splicingという現象も存在している。その場合、離反したスプライシング部位間でのスプライシングに伴って、除去される二つのイントロン間に存在している、一つまたは複数のエキソン領域も除去され、これら除去されるエキソン領域にコードされる部分アミノ酸配列は、得られるmRNA中には、コードされていないものとなる。また、前後のエキソン領域の連結部に跨って、コードされるアミノ酸自体は、N末端から同じ位置に存在するが、第三文字、あるいは、第二文字と第三文字が相違する結果、異なるアミノ酸となっている可能性もある。例えば、AG/TでコードされているSerが、AG/AでコードされているArgとなっていることもある。

## 【0045】

更には、alternative splicingは存在していないものの、ゲノム遺伝子解析において、仮に同定されたエキソン領域末端の連結に誤りがあり、先のエキソン末尾のACと後のエキソン先端のAとからなる、AC/AでコードされるThrと同定されていた結果が、先のエキソン末尾のAと後のエキソン先端のAAとからなる、A/AAでコードされるLysと同定すべきものであるという、同定上の誤謬がデータベース中に存在する可能性も排除できない。また、ゲノム遺伝子解析において、エキソン領域に関して、仮に同定はなされているものの、対応するmRNA、または、そのcDNAの塩基配列解析による検証がなされていない場合も少なくない。その場合、仮に同定はなされているエキソン領域とは異なり、その前後に位置するイントロンと判断された領域を含め、異なる読み枠（フレーム・シフト）において、見出されるオープン・リーディング領域複数が、実際には、エキソン領域であったという、同定上の誤謬がデータベース中に存在することも排除できない。いずれの場合も、データベース上に収録されている、（推断）完全長アミノ酸配列と対比して、現実には翻訳されているペプチド鎖において、相当するエキソン領域に該当する部分は、相違した部分アミノ酸配列を有している。

## 【0046】

加えて、mRNAから翻訳された完全長アミノ酸配列を有するペプチド鎖に対して、そのペプチド鎖内部で、前後する部位間でペプチド鎖が連結され、その間のペプチド断片が結果的に除去される、cis型のプロテイン・スプライシング過程も、特殊なケースでは存在することが報告されている。このcis型のプロテイン・スプライシング過程においても、最終的に得られるタンパク質は、完全長アミノ酸配列と対比させると、部分的にアミノ酸配列が欠損しているが、前述のalternative splicingの過程では、アミノ酸配列の欠損は、エキソン単位で生じるのとは異なり、cis型のプロテイン・スプライシング過程に起因するアミノ酸配列の欠損は、エキソン領域とは相関を有していないものである。

## 【0047】

上述する、翻訳後、完全長アミノ酸配列を有するペプチド鎖のN末端に存在するシグナル・ペプチド部などがプロセッシング過程で切除され、成熟型タンパク質へとなるもの以外に、例えば、一旦、N末側に、プレ配列やプロ配列を含むプレタンパク質、プロタンパク質として生合成され、その後、該プレ配列やプロ配列部の切除を受け、活性型のタンパク質へと変換されることも、数多く報告されている。また、活性型のタンパク質へと変換に際して、C末側のペプチド部分の切除がなされ、C末切除型タンパク質となる事例も、数多く報告されている。これら、翻訳後、完全長アミノ酸配列を有するペプチド鎖から、最終的に、そのN末側、あるいはC末側の所定の部分ペプチド鎖が切除を受けたタンパク質においては、残されたペプチド鎖は、完全長アミノ酸配列のうち、所定の連続したアミノ酸配列部分から構成されている。

## 【0048】

ゲノム遺伝子においては、互いに、そのアミノ酸配列が高い相同性を有する、相同体タ



ンパク質をそれぞれコードする複数の遺伝子が存在することも知られている。例えば、対立遺伝子間、あるいは、複対立遺伝子間において、相互にコードされるタンパク質のアミノ酸配列間に、極く僅かな相違が存在しており、対立相同体、複対立相同体タンパク質として、報告されている事例も、多数存在している。このような、互いに相長的ではあるものの、それぞれ異なる遺伝子によりそのアミノ酸配列がコードされているもの以外に、本来、同じ遺伝子座を示す遺伝子であるものが、その個体毎の多型性を反映して、その塩基配列に、極く僅かな相違が存在している遺伝子変異の存在も、多数報告されている。中でも、塩基配列の極く僅かな相違が、塩基配列全体の塩基長、エキソン、イントロンの構成に関しては変化を及ぼさないが、一つの塩基が他の塩基へと変異されており、かかる一塩基の変異に伴い、エキソン中において、その変異コドンによりコードされるアミノ酸種に相違が生じる遺伝子多型も存在している。この種の遺伝子多型は、「一塩基多型」と呼ばれ、特に、翻訳されるアミノ酸配列において、アミノ酸置換が生じると、「一塩基多型」に起因する変異体タンパク質が生合成されることになる。

#### 【0049】

以上に述べた、ゲノム遺伝子上にコードされる完全長アミノ酸配列を有するペプチド鎖と対比した際、実際に見出されるタンパク質を構成するペプチド鎖のアミノ酸配列に相違を有する場合に加えて、該タンパク質を構成している、ペプチド鎖に含まれるアミノ酸側鎖に対して、種々の酵素タンパク質が翻訳後作用して、修飾基の導入がなされることも、多くのタンパク質において、報告がなされている。

#### 【0050】

この翻訳後修飾の代表的なものとして、リン酸化、メチル化、アセチル化、ヒドロキシ化、ホルミル化、ピログルタミル化を挙げることができる。

#### 【0051】

メチル化の例としては、メチル・トランスフェラーゼによるメチル基転移反応のため、翻訳後のタンパク質において、アミノ基に対するメチル基置換（N-メチル化）、ヒドロキシ基に対するメチル基置換（O-メチル化）、スルファニル基に対するメチル基置換（S-メチル化）がある。より具体的には、アミノ酸残基の側鎖上へのメチル基としては、N-メチル化は、ヒスチジン、リシン、アルギニンなど、O-メチル化は、グルタミン酸、アスパラギン酸、S-メチル化は、システインに対して生じる。

#### 【0052】

また、リン酸化の例としては、タンパク質キナーゼによるリン酸化には、セリン／トレオニン・キナーゼが関与するセリン／トレオニン側鎖上のヒドロキシ基へのリン酸化、チロシン・キナーゼが関与するチロシン側鎖上のヒドロキシ基へのリン酸化、ホルミル化の例としては、ホルミル・トランスフェラーゼによる、N-ホルミルグルタミン酸、N-ホルミルメチオニンなど、アセチル化の例としては、アセチル化酵素による、N-アセチル化リシンなど、また、ヒドロキシ化の例としては、ヒドロキシラーゼによる、ヒドロキシプリン、5-ヒドロキシリシンなどを示すことができる。

#### 【0053】

上述するゲノム遺伝子上にコードされる完全長アミノ酸配列を有するペプチド鎖と対比した際、実際に見出されるタンパク質を構成するペプチド鎖のアミノ酸配列に相違を有する場合、ならびに、該タンパク質を構成している、ペプチド鎖に含まれるアミノ酸側鎖に対して、種々の酵素タンパク質が翻訳後作用して、修飾基の導入がなされている場合、当該タンパク質に対応するゲノム遺伝子上にコードされる完全長アミノ酸配列を有するペプチド鎖と対比して、実際のタンパク質を構成しているペプチド鎖は、それぞれの要因に起因する特異的な質量変化を示すものである。

#### 【0054】

本発明にかかる質量分析法を利用するタンパク質の同定方法は、特に、前述する解析対象タンパク質を構成するペプチド鎖は、対応するゲノム遺伝子上にコードされる完全長アミノ酸配列を有するペプチド鎖と対比して、種々の要因に起因する特異的な質量変化を示す場合においても、



解析対象タンパク質に対して、質量分析法で得られる情報に基づき、公知タンパク質に関するデータ・ベースに収録される、公知タンパク質個々に関して、該公知タンパク質を構成するペプチド鎖の完全長アミノ酸配列をコードしているゲノム遺伝子、および該完全長アミノ酸配列の翻訳を可能とする、mRNA中の読み枠の塩基配列、ならびに、その塩基配列でコードされる（推断）完全長アミノ酸配列の配列情報を参照して、該解析対象タンパク質に相当すると推定される、前記データ・ベースに収録される、公知タンパク質の一つを高い確度で選定する方法となる。すなわち、解析対象のタンパク質が、例えば、「翻訳後修飾」を有するタンパク質、あるいは、ある公知タンパク質に対する、スプライシング変異体、「一塩基多型」変異体などに相当する場合、一致の見られないペプチド断片をもたらす要因である、かかる「翻訳後修飾」やアミノ酸配列上の変異を解析する際、基準とすべき「公知タンパク質の候補」を高い蓋然性で特定することが可能な手段となる。

#### 【0055】

以下に、本発明にかかる質量分析法を利用するタンパク質の同定方法の原理、ならびに、前述する解析対象タンパク質を構成するペプチド鎖は、対応するゲノム遺伝子上にコードされる完全長アミノ酸配列を有するペプチド鎖と対比して、種々の要因に起因する特異的な質量変化を示している際、各要因に関して、本発明にかかる質量分析法を利用するタンパク質の同定方法を適用する具体的な形態について、より詳細に説明する。

#### 【0056】

(A) ゲノム遺伝子上にコードされる完全長アミノ酸配列を有するペプチド鎖からなるタンパク質の同定

本発明にかかる質量分析法を利用するタンパク質の同定方法においては、生体試料中に含まれる、解析対象のタンパク質を予め単離し、質量分析する際、未知の夾雑物に由来するイオン種が質量分析スペクトル中に出現することを回避している。

#### 【0057】

一方、一般に、単離した段階ではタンパク質は、その三次元的構造を保持している、また、ペプチド鎖内のシステイン・ブリッジ構造のような、Cys-Cys結合を有しているため、Cys-Cys結合中のジスルフィド(S-S)結合を開裂させることが可能な還元処理を施し、且つ、対象タンパク質の折り畳み(folding)を解き、対象タンパク質を構成するペプチド鎖を直鎖状にする処理を行っている。

#### 【0058】

以上の前処理を施した、直鎖状ペプチド鎖を分離し、さらに、特定のアミノ酸、またはアミノ酸配列において、選択的にペプチド鎖を切断する、部位特異的なタンパク質分解処理を施す。この部位特異的なタンパク質分解処理により、ペプチド鎖中に存在する、特異的な切断部位において、対象のペプチド鎖は断片化され、複数のペプチド断片が得られる。その際、ペプチド鎖上において、隣接している二つのペプチド断片について、一部は切断され、一部は切断を受けず連結されたままであると、その後の質量分析スペクトルの解釈を困難とする要因となる。従って、本発明の方法においては、対象タンパク質より採取される、直鎖状ペプチド鎖に由来するペプチド断片複数は、一部は切断され、一部は切断を受けず連結されたまま残ることがないように、概ね等量的かつ選択的に切断がなされたものに調製する。

#### 【0059】

すなわち、低分子量の有機化合物では、質量分析法を利用する、その構造解析においては、該有機化合物の親イオン種と、そのフラグメンテーション(断片化)により生成する各種の娘イオン種の分子量(M/Z)を測定し、その分子構造の推定を行うが、タンパク質は、質量分析法によって、その親イオン種の分子量を決定することは一般に困難であるので、解析対象のタンパク質から採取した直鎖状ペプチド鎖を、予め、等量的かつ選択的に断片化を施しておき、これらのペプチド断片複数全てについて、対応するペプチド断片「親イオン種」の分子量を測定することで、元の直鎖状ペプチド鎖に由来する娘イオン種の分子量として利用している。また、原理的には、各対応するペプチド断片「親イオン種

」の分子量を積算することにより、元の直鎖状ペプチド鎖が有する分子量を算出することができる。

#### 【0060】

その際、各ペプチド断片の「親イオン種」について、MS/MS分析を行うことで、そのフラグメンテーション（断片化）により生成する各種の娘イオン種の分子量を合わせて、測定することも可能である。場合によっては、ペプチド断片「親イオン種」の分子量と、その断片化により生成する各種の「娘イオン種」の分子量の情報を総合することで、各ペプチド断片中に含まれているアミノ酸残基の種類と個数を推定することが可能である場合も少なくない。しかしながら、各ペプチド断片自体、複数個のアミノ酸残基を含むペプチド鎖であるため、含まれているアミノ酸残基の種類と個数が推定されても、その連結順序、すなわち、部分的なアミノ酸配列全体を特定することは、一般に困難である。同様に、各ペプチド断片複数が、元の直鎖状ペプチド鎖において、どのような順序で連結されていたかを特定することも、一般に困難である。

#### 【0061】

そのため、本発明の方法では、解析対象のタンパク質は、既にそのアミノ酸配列に関する情報が報告されている公知のタンパク質と同一のもの、ないしは、該公知のタンパク質をコードしている遺伝子の産物と仮定した上で、その同定候補となる公知タンパク質を選択する、下記する手法を採用している。

#### 【0062】

仮に、公知タンパク質の一つが、解析対象のタンパク質と同一のアミノ酸配列を有するペプチド鎖で構成されているとすれば、その公知タンパク質に対して、同じく、直鎖状ペプチド鎖とし、部位特異的なタンパク質分解処理を施して、複数のペプチド断片とした際、質量分析において、各ペプチド断片の「親イオン種」の分子量に関しては、原理的に一致する結果を与える。但し、多くの種類の公知タンパク質について、その標準試料を実際に入手して、対比測定を行うことは、現実では、容易でないため、本発明では、公知タンパク質について、報告されている完全長アミノ酸配列を参照し、前記Cys側鎖上のスルファニル基（-SH）に対する還元処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、完全長アミノ酸配列を有するペプチド鎖に由来するペプチド断片複数を推定する。この推定ペプチド断片において、それが有するアミノ酸配列は、推定した時点で決定されており、対応する分子量を算定することができる。公知タンパク質の標準試料における、各ペプチド断片「親イオン種」の分子量実測値のセットに代えて、該公知タンパク質に対する（推断）完全長アミノ酸配列に基づき、前記の要領で推定される、各公知タンパク質に由来する、推定ペプチド断片複数が有する推定分子量（Mref）のセットを利用する。

#### 【0063】

本発明の方法で利用する、公知タンパク質に関するデータ・ベースに収録される、公知タンパク質個々に関して、該公知タンパク質を構成するペプチド鎖の完全長アミノ酸配列をコードしているゲノム遺伝子、および該完全長アミノ酸配列への翻訳を可能とする、mRNA中の読み枠の塩基配列、ならびに、その塩基配列でコードされる（推断）完全長アミノ酸配列の配列情報を参照して、前記の要領に従って、データ・ベース中に収録される、各公知タンパク質について、該公知タンパク質に由来する、推定ペプチド断片複数が有する推定分子量（Mref）のセットを予め作成した上で、該公知タンパク質に由来する、推定ペプチド断片複数が有する推定分子量（Mref）のセット全数から構成される、ペプチド断片複数の推定分子量（Mref）のデータ・セットを、参照基準データ・ベースとして利用する。

#### 【0064】

解析対象タンパク質については、少なくとも、生成する前記ペプチド断片複数の質量（M）について、対応する一価の「親の陽イオン種」の分子量（ $M+H/Z$ ； $Z=1$ ）、または一価の「親の陰イオン種」の分子量（ $M-H/Z$ ； $Z=1$ ）として、質量分析法で測定される結果に基づき、決定されるペプチド断片複数の各質量実測値（Mex）のセットを



用意する。また、付加的に、MS/MS分析法によって、各ペプチド断片に対応する一価の「親の陽イオン種」、または一価の「親の陰イオン種」について、その断片化により生成する各種の娘イオン種の分子量の測定結果を、二次的質量分析結果として、取得する。

#### 【0065】

一次対比操作においては、まず、

解析対象タンパク質において決定されるペプチド断片複数の各質量実測値 (Mex) のセットと、

参照基準データ・ベース中の、各公知タンパク質に由来する、推定ペプチド断片複数の有する推定分子量 (Mref) のセットとを対比させ、

利用している質量分析法自体に起因する測定誤差を考慮して、各質量実測値 (Mex) と、各公知タンパク質に由来するセット中の、推定ペプチド断片複数の有する推定分子量 (Mref) とが実質的に一致すると判定される、該解析対象タンパク質由来の実測ペプチド断片数 (Nex-id) ならびに該公知タンパク質由来の推定ペプチド断片数 (Nref-id) を決定する。

#### 【0066】

なお、場合によっては、公知タンパク質由来の推定ペプチド断片中に、推定分子量 (Mref) が等しい、あるいは、分子量差が1しか異なっていない、極めて類似する推定分子量 (Mref) を有する、複数の推定ペプチド断片が偶々存在することもある。その際、該解析対象タンパク質由来の実測ペプチド断片の質量実測値 (Mex) は、測定誤差の範囲内で、これら複数の推定ペプチド断片の推定分子量 (Mref) のいずれとも、実質的に一致していると思わせることもある。このように、一意的な「一致の判定」が困難な場合には、前記の二次的質量分析結果、例えば、MS/MS分析法で得られる、各種の娘イオン種の分子量の測定結果、ピーク強度、ピークの半値幅などをも参照して、複数種の実測ペプチド断片ピークが、見掛け上一つのピークを形成しているか、否か、また、何種のピークが重なっているかを判別することもできる。最終的には、種々の因子を考慮して、一意的な「一致の判定」が困難な場合にも、前記複数の推定ペプチド断片のいずれと、実測ペプチド断片とが一致すると判定されるか、統計的な確率の重み付けをした上で、「一致の判定」を行う、公知タンパク質由来の推定ペプチド断片を選別する。かかる統計的な確率の重み付けは、一意的な「一致の判定」が可能である場合には、確率1、二次的質量分析結果を参照しても、なお、二種の推定ペプチド断片のいずれか判別が困難である際には、確率1/2となる。前述する、一致すると判定される、該解析対象タンパク質由来の実測ペプチド断片数 (Nex-id) ならびに該公知タンパク質由来の推定ペプチド断片数 (Nref-id) を決定する際には、かかる統計的な確率の重み付けを付して、その一致断片数を算出することになる。

#### 【0067】

この一次対比操作において決定される、各公知タンパク質に対して、一致すると判定される、該解析対象タンパク質由来の実測ペプチド断片数 (Nex-id) ならびに該公知タンパク質由来の推定ペプチド断片数 (Nref-id) の多い順に、公知タンパク質を選択し、その一致数 (Nex-id=Nref-id) が最大を示す公知タンパク質を選択して、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質の群とする。

#### 【0068】

前記参照基準データ・ベース中に含まれる公知タンパク質の一つが、解析対象のタンパク質と同一のアミノ酸配列を有するペプチド鎖で構成されているとすれば、少なくとも、前記一次対比操作において、選別される、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群の中に、当然、かかる解析対象のタンパク質と同一のアミノ酸配列を有するペプチド鎖で構成されている公知タンパク質が含まれることになる。また、該解析対象タンパク質由来の実測ペプチド断片複数全ては、該公知タンパク質由来の推定ペプチド断片の推定分子量 (Mref) とが実質的に一致すると判定されるはずである。その際、多くの場合、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群は、この解析対象のタンパク質と同一のアミノ酸配列を有するペプチド



鎖で構成されている公知タンパク質のみとなる。換言するならば、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群に含まれる公知タンパク質の種類が一つである場合、このデータ・ベース中より選別される該一種類の公知タンパク質を、該解析対象タンパク質に対する、単一の同定の候補と判定することができる。

#### 【0069】

偶々、公知タンパク質二種以上が、推定ペプチド断片複数が有する推定分子量 (Mref) のセットとしては、完全に同じ値を有することも、排除できない。

#### 【0070】

従って、解析対象のタンパク質由来のペプチド断片各質量実測値 (Mex) と、公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) とが実質的に一致すると判定される場合、MS/MS分析法によって得られる、解析対象のタンパク質由来の各ペプチド断片に対応する一価の「親の陽イオン種」、または一価の「親の陰イオン種」について、その断片化により生成する各種の娘イオン種の分子量の測定結果と、ペプチド断片の分子量が一致すると判定される、公知タンパク質に由来する推定ペプチド断片が有するアミノ酸配列から、MS/MS分析において、その断片化により生成すると想定される各種の娘イオン種の分子量推定値との間において、その対応を確認することにより、より確度の高い判定が可能となる。

#### 【0071】

具体的には、解析対象のタンパク質由来の各ペプチド断片に対応する一価の「親の陽イオン種」、または一価の「親の陰イオン種」について、その断片化により生成する各種の娘イオン種の分子量の測定結果は、例えば、該ペプチド断片に含まれる、部分ペプチド鎖に相当する娘イオン種の分子量を示すことがあり、公知タンパク質に由来する推定ペプチド断片が有するアミノ酸配列から、対応する娘イオン種が生成するか否かを確認することにより、偶々、公知タンパク質二種以上が、推定ペプチド断片複数が有する推定分子量 (Mref) のセットとしては、完全に同じ値を有する際も、かかる二次的な質量分析結果を利用することで、より確度の高い単一の同定の候補を、選別することができる。

#### 【0072】

さらには、ペプチド鎖C末端の部分的アミノ酸配列に関しては、例えば、国際公開 WO 03/081255 A1号パンフレットに開示される「ペプチドのC末端アミノ酸配列解析方法」の手法などを利用することで、質量分析法によって、少なくとも、数個のアミノ酸について、そのペプチド鎖のC末端側のアミノ酸配列を特定することが可能である。高い確度で解析が可能となっている。MS/MS分析法によって得られる、解析対象のタンパク質由来の各ペプチド断片に対応する一価の「親の陽イオン種」、または一価の「親の陰イオン種」について、その断片化により生成する各種の娘イオン種の分子量の測定結果に代えて、あるいは、さらに付加的に、前記の「ペプチドのC末端アミノ酸配列解析方法」の手法を利用して入手される、解析対象のタンパク質由来の各ペプチド断片におけるC末端側のアミノ酸配列情報をも、二次的な質量分析結果として、利用することで、公知タンパク質に由来する推定ペプチド断片が有するアミノ酸配列との部分的な一致も確認でき、より確度の高い単一の同定の候補を、選別することができる。

#### 【0073】

(B) 翻訳後修飾を有するペプチド鎖からなるタンパク質の同定

解析対象のタンパク質は、ゲノム遺伝子上にコードされる完全長アミノ酸配列を有するペプチド鎖からなるタンパク質ではあるが、そのペプチド鎖上に翻訳後修飾を有しているタンパク質であると想定する。

#### 【0074】

その場合、上述する前処理済みの直鎖状ペプチド鎖とし、部位特異的なタンパク質分解処理を施して、複数のペプチド断片とした際、質量分析において、各ペプチド断片の「親イオン種」の分子量に関しては、かかる翻訳後修飾がなされているアミノ酸残基を含むペプチド断片の「親イオン種」の分子量は、翻訳後修飾が無い対応のペプチド断片の「親イオン種」の分子量と相違したものとなる。

## 【0075】

なお、翻訳後修飾の代表的なものとして、リン酸化、メチル化、アセチル化、ヒドロキシ化、ホルミル化、ピログルタミル化を挙げることができる。より具体的には、N-メチル化は、ヒスチジン、リシン、アルギニンなど、O-メチル化は、グルタミン酸、アスパラギン酸、S-メチル化は、システインに対して生じる。また、リン酸化は、セリン/トアラニン側鎖上のヒドロキシ基へのリン酸化、チロシン側鎖上のヒドロキシ基へのリン酸化が、ホルミル化は、N-ホルミルグルタミン酸、N-ホルミルメチオニンなど、アセチル化の例としては、アセチル化酵素による、N-アセチル化リシンなど、また、ヒドロキシ化の例としては、ヒドロキシプロリン、5-ヒドロキシリシンなどが想定できる。

## 【0076】

仮に、前記参照基準データ・ベース中に含まれる公知タンパク質の一つが、解析対象のタンパク質と同一のアミノ酸配列を有するペプチド鎖で構成されているとすれば、上述する一次対比操作において、少なくとも、かかる公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と実質的に一致すると判定される、該解析対象タンパク質由来の実測ペプチド断片数 (Nex-id) ならびに該公知タンパク質由来の推定ペプチド断片数 (Nref-id) を決定すると、原理的には、該解析対象のタンパク質に由来する翻訳後修飾がなされているアミノ酸残基を含むペプチド断片数 (Nex-mod) を、該解析対象タンパク質由来の実測ペプチド断片総数 (Nex) から減じた数となる。

## 【0077】

該解析対象のタンパク質に由来する翻訳後修飾がなされているアミノ酸残基を含むペプチド断片の分子量と、偶々、同じ分子量を示す、アミノ酸配列を有し、翻訳後修飾が存在しないペプチド断片の存在は、完全には、排除できないものの、そのような確率は相当に低い。

## 【0078】

従って、前記参照基準データ・ベース中に含まれる公知タンパク質の一つが、解析対象のタンパク質と同一のアミノ酸配列を有するペプチド鎖で構成されているとすれば、少なくとも、前記一次対比操作において、選別される、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群の中に、極めて高い確率で、かかる解析対象のタンパク質と同一のアミノ酸配列を有するペプチド鎖で構成されている公知タンパク質が含まれることになる。その際、相当に高い確率で、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群は、この解析対象のタンパク質と同一のアミノ酸配列を有するペプチド鎖で構成されている公知タンパク質のみとなる。換言するならば、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群に含まれる公知タンパク質種類が一つである場合、このデータ・ベース中より選別される該種類の公知タンパク質を、該解析対象タンパク質に対する、単一の同定の候補と判定することができる。

## 【0079】

前記 (A) の事例と同様に、解析対象のタンパク質由来のペプチド断片各質量実測値 (Mex) と、公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) とが実質的に一致すると判定される場合、MS/MS分析法によって得られる、解析対象のタンパク質由来の各ペプチド断片に対応する一価の「親の陽イオン種」、または一価の「親の陰イオン種」について、その断片化により生成する各種の娘イオン種の分子量の測定結果と、ペプチド断片の分子量が一致すると判定される、公知タンパク質に由来する推定ペプチド断片が有するアミノ酸配列から、MS/MS分析において、その断片化により生成すると想定される各種の娘イオン種の分子量推定値との間において、その対応を確認することにより、より確度の高い判定が可能となる。さらには、MS/M S分析法によって得られる、解析対象のタンパク質由来の各ペプチド断片に対応する一価の「親の陽イオン種」、または一価の「親の陰イオン種」について、その断片化により生成する各種の娘イオン種の分子量の測定結果に代えて、あるいは、さらに付加的に、前記の「ペプチドのC末端アミノ酸配列解析方法」の手法を利用して入手される、解析対象の



タンパク質由来の各ペプチド断片におけるC末端側のアミノ酸配列情報をも、二次的な質量分析結果として、利用することで、公知タンパク質に由来する推定ペプチド断片が有するアミノ酸配列との部分的な一致も確認でき、より確度の高い単一の同定の候補を、選別することができる。

**【0080】**

また、単一の同定の候補として選択された公知タンパク質由来の推定ペプチド断片のうち、一次対比操作において、該解析対象タンパク質由来実測ペプチド断片と分子量が一致すると判定されていない、未同定の推定ペプチド断片について、そのペプチド鎖上に、翻訳後修飾が生じる可能性を持つアミノ酸残基が存在している場合、そのアミノ酸残基の側鎖への修飾基付加に起因する翻訳後修飾が存在すると仮定した際に予測される、アミノ酸残基の側鎖への修飾基付加に起因する翻訳後修飾を有する推定ペプチド断片に対する推定分子量 (Mref) を改めて算定する。

**【0081】**

次いで、この修飾基付加に起因する翻訳後修飾を有する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

修飾基付加に起因する翻訳後修飾を有する推定ペプチド断片に対する推定分子量 (Mref) と一致する、質量実測値 (Mex) を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、一次対比操作の結果に基づき、選別されている、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

**【0082】**

また、この二次対比操作において、修飾基付加に起因する翻訳後修飾を有する推定ペプチド断片に対する推定分子量 (Mref) と一致すると判定された、該解析対象タンパク質由来実測ペプチド断片に関しても、MS/MS分析法によって得られる、解析対象のタンパク質由来の各ペプチド断片に対応する一価の「親の陽イオン種」、または一価の「親の陰イオン種」について、その断片化により生成する各種の娘イオン種の分子量の測定結果と、ペプチド断片の分子量が一致すると判定される、公知タンパク質に由来する推定ペプチド断片が有するアミノ酸配列から、MS/MS分析において、その断片化により生成すると想定される各種の娘イオン種の分子量推定値との間において、その対応を確認することにより、より確度の高い判定が可能となる。さらには、MS/MS分析法によって得られる、解析対象のタンパク質由来の各ペプチド断片に対応する一価の「親の陽イオン種」、または一価の「親の陰イオン種」について、その断片化により生成する各種の娘イオン種の分子量の測定結果に代えて、あるいは、さらに付加的に、前記の「ペプチドのC末端アミノ酸配列解析方法」の手法を利用して入手される、解析対象のタンパク質由来の各ペプチド断片におけるC末端側のアミノ酸配列情報をも、二次的な質量分析結果として、利用することで、公知タンパク質に由来する推定ペプチド断片が有するアミノ酸配列との部分的な一致も確認でき、より確度の高い単一の同定の候補を、選別することができる。

**【0083】**

例えば、ホルミル化のうち、N-ホルミルメチオニン、メチオニン-tRNA ホルミル・トランスフェラーゼの作用により、N-ホルミルメチオニン-tRNAとして合成され、ペプチド鎖への翻訳に際し、N末端のメチオニンに代えて、導入されることも多く、このN末端のN-ホルミルメチオニンによる、修飾を受けている解析対象タンパク質では、かかるN末のペプチド断片を除き、それ以降のペプチド断片に関しては、いずれも、該解析対象タンパク質由来の実測ペプチド断片は、単一の同定の候補である公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と実質的に一致すると判定される。

**【0084】**

その際には、一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質の配列情報を参照し、



一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致すると判定されている、該解析対象タンパク質由来の実測ペプチド断片複数について、対応している該公知タンパク質に由来する推定ペプチド断片が占めるべき位置に配置した際、一致すると判定される前記実測ペプチド断片の一群が、該公知タンパク質の完全長アミノ酸配列中に含まれる、連続したアミノ酸配列を構成している場合、

かかる一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0085】

勿論、二次対比操作において、該同定の候補と判定される公知タンパク質に由来する、修飾基付加に起因する翻訳後修飾を有する推定ペプチド断片が有する推定分子量 (Mref-mod) と、一致すると判定されている、該解析対象タンパク質由来の実測ペプチド断片を含めて、対応している該公知タンパク質に由来する推定ペプチド断片が占めるべき位置に配置した際、一致すると判定される前記実測ペプチド断片の一群が、該公知タンパク質の完全長アミノ酸配列中に含まれる、連続したアミノ酸配列を構成している場合、

かかる一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0086】

##### (C) N末切除型タンパク質の同定

解析対象のタンパク質は、ゲノム遺伝子上にコードされる完全長アミノ酸配列を有するペプチド鎖として翻訳された後、そのN末端に存在する、シグナル・ペプチド部が切除された成熟型タンパク質、あるいは、プレ配列あるいはプロ配列部が切除を受け、活性型となったタンパク質など、N末切除型タンパク質であると想定する。

#### 【0087】

その場合、上述する前処理済みの直鎖状ペプチド鎖とし、部位特異的なタンパク質分解処理を施して、複数のペプチド断片とした際、質量分析において、各ペプチド断片の「親イオン種」の分子量に関しては、切除を受けているN末部分に含まれるペプチド断片は、元々存在してなく、さらには、かかるN末端部の切除が生じている部分アミノ酸配列を含むペプチド断片の「親イオン種」の分子量は、N末端部の切除が生じていない場合の対応するペプチド断片の「親イオン種」の分子量と相違したものとなる。具体的には、そのペプチド鎖においては、N末側が短縮を受けており、分子量がより小さいものとなっている。

#### 【0088】

仮に、前記参照基準データ・ベース中に含まれる公知タンパク質の一つに対する（推定）完全長アミノ酸配列が、解析対象のタンパク質の完全長アミノ酸配列と同一のアミノ酸配列を有するとすれば、上述する一次対比操作において、少なくとも、かかる公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と実質的に一致すると判定される、該解析対象タンパク質由来の実測ペプチド断片数 (Nex-id) ならびに該公知タンパク質由来の推定ペプチド断片数 (Nref-id) を決定すると、原理的には、該解析対象のタンパク質に由来するN末端のペプチド断片を除き、一致すると判定されるため、該解析対象タンパク質由来の実測ペプチド断片総数 (Nex) から1を減じた数となる。

#### 【0089】

該解析対象のタンパク質に由来するN末端のペプチド断片の分子量と、偶々、同じ分子量を示す、推定ペプチド断片を有し、残る該解析対象タンパク質由来の実測ペプチド断片数 (Nex-1) に関しても、それらの質量実測値 (Mex) と一致する、推定ペプチド断片複数が有する推定分子量 (Mref) を示す、別種の公知タンパク質の存在は、完全には、排除できないものの、そのような確率は相当に低い。

## 【0090】

従って、前記参照基準データ・ベース中に含まれる公知タンパク質の一つに対する（推断）完全長アミノ酸配列が、解析対象のタンパク質の完全長アミノ酸配列と同一のアミノ酸配列を有するとすれば、少なくとも、前記一次対比操作において、選別される、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群の中に、極めて高い確率で、かかる解析対象のタンパク質の完全長アミノ酸配列と同一のアミノ酸配列を有する、（推断）完全長アミノ酸配列を有する公知タンパク質が含まれることになる。その際、相当に高い確率で、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群は、この解析対象のタンパク質の完全長アミノ酸配列と同一のアミノ酸配列を有する、（推断）完全長アミノ酸配列を有する公知タンパク質のみとなる。換言するならば、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群に含まれる公知タンパク質種類が一つである場合、このデータ・ベース中より選別される該一種類の公知タンパク質を、該解析対象タンパク質に対する、単一の同定の候補と判定することができる。

## 【0091】

また、単一の同定の候補として選別された公知タンパク質に対する（推断）完全長アミノ酸配列が、解析対象のタンパク質の完全長アミノ酸配列と同一のアミノ酸配列を有する際には、一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される該公知タンパク質の配列情報を参照し、

一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量（Mref）と、一致すると判定されている、該解析対象タンパク質由来の実測ペプチド断片複数について、対応している該公知タンパク質に由来する推定ペプチド断片が占めるべき位置に配置した際、該解析対象タンパク質由来のペプチド断片が全て検出されている場合には、一致すると判定される前記実測ペプチド断片の一群が、該公知タンパク質の完全長アミノ酸配列中に含まれる、連続したアミノ酸配列を構成しているはずであり、すなわち、N末側を除き、該公知タンパク質の完全長アミノ酸配列中のC末端へと続く、連続したアミノ酸配列を構成しているはずであり

、かかる一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

## 【0092】

加えて、該解析対象タンパク質由来のペプチド断片が全て検出されている場合には、前記一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量（Mref）と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が唯一つ残っており、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分と連結される、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定のN末側に存在する推定ペプチド断片の群部分について、翻訳後、そのN末側の一部を切断されるプロセッシングが生じ、成熟型タンパク質へと変換されると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、翻訳後のN末プロセッシングに由来する一連の推定ペプチド断片複数の推定分子量（Mref）を算出し、

前記翻訳後のN末プロセッシングに由来する一連の推定ペプチド断片に対する推定分子量（Mref）のうち、唯一残されている未同定の該解析対象タンパク質由来実測ペプチド断片の質量実測値（Mex）と一致するものの有無を判定する、二次対比操作を行うと、

前記翻訳後のN末プロセッシングに由来する一連の推定ペプチド断片に対する推定分子量（Mref）のうち、唯一残されている未同定の該解析対象タンパク質由来実測ペプチド断片の質量実測値（Mex）と一致するものが一つ選別されるはずである。実際に、かかる



二次対比操作において、一致するものの存在が検証された場合、上記の一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

**【0093】**

場合によっては、該解析対象タンパク質由来のペプチド断片の全ては、検出されていないこともある。その際にも、前記一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が唯一つ残っているのみとなるはずである。一方、一致すると判定されている、該解析対象タンパク質由来の実測ペプチド断片複数について、対応している該公知タンパク質に由来する推定ペプチド断片が占めるべき位置に配置した際、一致すると判定される前記実測ペプチド断片の一群は、該公知タンパク質の完全長アミノ酸配列中に含まれる、N末側を除き、検出されていないペプチド断片に由来する、未同定領域があるものの、該公知タンパク質の完全長アミノ酸配列中のC末端へと続く、連続したアミノ酸配列を構成しているはずである。また、唯一つ残っている、未同定の該解析対象タンパク質由来実測ペプチド断片に関して、同様の二次対比操作を行うと、前記翻訳後のN末プロセッシングに由来する一連の推定ペプチド断片に対する推定分子量 (Mref) のうち、唯一残されている未同定の該解析対象タンパク質由来実測ペプチド断片の質量実測値 (Mex) と一致するものが一つ選別されるはずである。実際に、かかる二次対比操作において、一致するものの存在が検証された場合、上記の一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

**【0094】**

勿論、この二次対比操作において、翻訳後のN末プロセッシングに由来する一連の推定ペプチド断片に対する推定分子量 (Mref) の一つと一致すると判定された、該解析対象タンパク質由来実測ペプチド断片に対しても、MS/MS分析法によって得られる、解析対象のタンパク質由来の各ペプチド断片に対応する一価の「親の陽イオン種」、または一価の「親の陰イオン種」について、その断片化により生成する各種の娘イオン種の分子量の測定結果と、ペプチド断片の分子量が一致すると判定される、公知タンパク質に由来する推定ペプチド断片が有するアミノ酸配列から、MS/MS分析において、その断片化により生成すると想定される各種の娘イオン種の分子量推定値との間において、その対応を確認することにより、より確度の高い判定が可能となる。さらには、MS/MS分析法によって得られる、解析対象のタンパク質由来の各ペプチド断片に対応する一価の「親の陽イオン種」、または一価の「親の陰イオン種」について、その断片化により生成する各種の娘イオン種の分子量の測定結果に代えて、あるいは、さらに付加的に、前記の「ペプチドのC末端アミノ酸配列解析方法」の手法を利用して入手される、解析対象のタンパク質由来の各ペプチド断片におけるC末端側のアミノ酸配列情報をも、二次的な質量分析結果として、利用することで、公知タンパク質に由来する推定ペプチド断片が有するアミノ酸配列との部分的な一致も確認でき、より確度の高い単一の同定の候補を、選別することができる。

**【0095】**

偶然、翻訳後のN末プロセッシングを生じさせる、エンドペプチダーゼによる切断部位と、部位特異的なタンパク質分解処理による切断が起こる部位とが一致することもある。その場合、結果的に、一次対比操作の際、未同定の該解析対象タンパク質由来実測ペプチド断片が残らないことになり、かかる場合も、当然、かかる一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

**【0096】**

(D) C末切除型タンパク質の同定

図3に例示するように、解析対象のタンパク質は、ゲノム遺伝子上にコードされる完全



長アミノ酸配列を有するペプチド鎖として翻訳された後、そのC末端の部分ペプチド鎖が切除を受け、活性型となったタンパク質など、C末切除型タンパク質であると想定する。

#### 【0097】

その場合、上述する前処理済みの直鎖状ペプチド鎖とし、部位特異的なタンパク質分解処理を施して、複数のペプチド断片とした際、質量分析において、各ペプチド断片の「親イオン種」の分子量に関しては、切除を受けているC末部分に含まれるペプチド断片は、元々存在してなく、さらには、かかるC末端部の切除が生じている部分アミノ酸配列を含むペプチド断片の「親イオン種」の分子量は、C末端部の切除が生じていない場合の対応するペプチド断片の「親イオン種」の分子量と相違したものとなる。具体的には、そのペプチド鎖では、C末側が短縮を受けており、分子量がより小さいものとなっている。

#### 【0098】

仮に、前記参照基準データ・ベース中に含まれる公知タンパク質の一つに対する（推断）完全長アミノ酸配列が、解析対象のタンパク質の完全長アミノ酸配列と同一のアミノ酸配列を有するとすれば、上述する一次対比操作において、少なくとも、かかる公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量（Mref）と実質的に一致すると判定される、該解析対象タンパク質由来の実測ペプチド断片数（Nex-id）ならびに該公知タンパク質由来の推定ペプチド断片数（Nref-id）を決定すると、該解析対象のタンパク質に由来するC末端のペプチド断片を除き、一致すると判定されるため、該解析対象タンパク質由来の実測ペプチド断片総数（Nex）から1を減じた数となる。

#### 【0099】

該解析対象のタンパク質に由来するN末端のペプチド断片の分子量と、偶々、同じ分子量を示す、推定ペプチド断片を有し、残る該解析対象タンパク質由来の実測ペプチド断片数（Nex-1）に関しても、それらの質量実測値（Mex）と一致する、推定ペプチド断片複数が有する推定分子量（Mref）を示す、別種の公知タンパク質の存在は、完全には、排除できないものの、そのような確率は相当に低い。

#### 【0100】

従って、前記参照基準データ・ベース中に含まれる公知タンパク質の一つに対する（推断）完全長アミノ酸配列が、解析対象のタンパク質の完全長アミノ酸配列と同一のアミノ酸配列を有するとすれば、少なくとも、前記一次対比操作において、選別される、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群の中に、極めて高い確率で、かかる解析対象のタンパク質の完全長アミノ酸配列と同一のアミノ酸配列を有する、（推断）完全長アミノ酸配列を有する公知タンパク質が含まれることになる。その際、相当に高い確率で、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群は、この解析対象のタンパク質の完全長アミノ酸配列と同一のアミノ酸配列を有する、（推断）完全長アミノ酸配列を有する公知タンパク質のみとなる。換言するならば、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群に含まれる公知タンパク質種類が一つである場合、このデータ・ベース中より選別される該一種類の公知タンパク質を、該解析対象タンパク質に対する、単一の同定の候補と判定することができる。

#### 【0101】

また、単一の同定の候補として選別された公知タンパク質に対する（推断）完全長アミノ酸配列が、解析対象のタンパク質の完全長アミノ酸配列と同一のアミノ酸配列を有する際には、一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される該公知タンパク質の配列情報を参照し、

一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量（Mref）と、一致すると判定されている、該解析対象タンパク質由来の実測ペプチド断片複数について、対応している該公知タンパク質に由来する推定ペプチド断片が占めるべき位置に配置した際、該解析対象タンパク質由来のペプチド断片が全て検出されている場合には、一致すると判定される前記実測ペプチド断片の一群が、該公知タンパク質の完全長アミノ酸配列中に含まれる、連続したア

ミノ酸配列を構成しているはずであり、すなわち、C末側を除き、該公知タンパク質の完全長アミノ酸配列中のN末端から続く、連続したアミノ酸配列を構成しているはずであり、

かかる一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0102】

加えて、該解析対象タンパク質由来のペプチド断片が全て検出されている場合には、前記一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が唯一つ残っており、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分と連結される、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定のC末側に存在する推定ペプチド断片の群部分について、翻訳後、そのC末側の一部を切断されるプロセッシングが生じ、C末切除型タンパク質へと変換されると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、翻訳後のC末プロセッシングに由来する一連の推定ペプチド断片複数の推定分子量 (Mref) を算出し、

前記翻訳後のC末プロセッシングに由来する一連の推定ペプチド断片に対する推定分子量 (Mref) のうち、唯一残されている未同定の該解析対象タンパク質由来実測ペプチド断片の質量実測値 (Mex) と一致するものの有無を判定する、二次対比操作を行うと、

前記翻訳後のC末プロセッシングに由来する一連の推定ペプチド断片に対する推定分子量 (Mref) のうち、唯一残されている未同定の該解析対象タンパク質由来実測ペプチド断片の質量実測値 (Mex) と一致するものが一つ選別されるはずである。実際に、かかる二次対比操作において、一致するものの存在が検証された場合、上記の一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0103】

場合によっては、該解析対象タンパク質由来のペプチド断片の全ては、検出されていないこともある。その際にも、前記一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が唯一つ残っているのみとなるはずである。一方、一致すると判定されている、該解析対象タンパク質由来の実測ペプチド断片複数について、対応している該公知タンパク質に由来する推定ペプチド断片が占めるべき位置に配置した際、一致すると判定される前記実測ペプチド断片の一群は、該公知タンパク質の完全長アミノ酸配列中に含まれる、C末側を除き、検出されていないペプチド断片に由来する、未同定領域があるものの、該公知タンパク質の完全長アミノ酸配列中のN末端から続く、連続したアミノ酸配列を構成しているはずである。また、唯一つ残っている、未同定の該解析対象タンパク質由来実測ペプチド断片に関して、同様の二次対比操作を行うと、前記翻訳後のC末プロセッシングに由来する一連の推定ペプチド断片に対する推定分子量 (Mref) のうち、唯一残されている未同定の該解析対象タンパク質由来実測ペプチド断片の質量実測値 (Mex) と一致するものが一つ選別されるはずである。実際に、かかる二次対比操作において、一致するものの存在が検証された場合、上記の一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0104】

勿論、この二次対比操作において、翻訳後のC末プロセッシングに由来する一連の推定ペプチド断片に対する推定分子量 (Mref) の一つと一致すると判定された、該解析対象



タンパク質由来実測ペプチド断片に関しても、MS/MS分析法によって得られる、解析対象のタンパク質由来の各ペプチド断片に対応する一価の「親の陽イオン種」、または一価の「親の陰イオン種」について、その断片化により生成する各種の娘イオン種の分子量の測定結果と、ペプチド断片の分子量が一致すると判定される、公知タンパク質に由来する推定ペプチド断片が有するアミノ酸配列から、MS/MS分析において、その断片化により生成すると想定される各種の娘イオン種の分子量推定値との間において、その対応を確認することにより、より確度の高い判定が可能となる。さらには、MS/MS分析法によって得られる、解析対象のタンパク質由来の各ペプチド断片に対応する一価の「親の陽イオン種」、または一価の「親の陰イオン種」について、その断片化により生成する各種の娘イオン種の分子量の測定結果に代えて、あるいは、さらに付加的に、前記の「ペプチドのC末端アミノ酸配列解析方法」の手法を利用して入手される、解析対象のタンパク質由来の各ペプチド断片におけるC末端側のアミノ酸配列情報をも、二次的な質量分析結果として、利用することで、公知タンパク質に由来する推定ペプチド断片が有するアミノ酸配列との部分的な一致も確認でき、より確度の高い単一の同定の候補を、選別することができる。

**【0105】**

また、前記の「ペプチドのC末端アミノ酸配列解析方法」の手法を利用して、該解析対象のタンパク質自体について、そのC末端側のアミノ酸配列情報を入手できる場合には、先に二次対比操作において、唯一残されている未同定の該解析対象タンパク質由来実測ペプチド断片の質量実測値 (Mex) と一致するものとして、一つ選別された、翻訳後のC末プロセッシングに由来する推定ペプチド断片のアミノ酸配列と対比を行うことで、二次対比操作の妥当性を検証することができる。

**【0106】**

偶然、翻訳後のC末プロセッシングを生じさせる、エンドペプチダーゼによる切断部位と、部位特異的なタンパク質分解処理による切断が起こる部位とが一致することもある。その場合、結果的に、一次対比操作の際、未同定の該解析対象タンパク質由来実測ペプチド断片が残らないことになり、かかる場合も、当然、かかる一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

**【0107】**

(E) プロテイン・スプライシングにより生成されるタンパク質の同定  
図2に例示するように、解析対象のタンパク質は、ゲノム遺伝子上にコードされる完全長アミノ酸配列を有するペプチド鎖として翻訳された後、そのペプチド鎖の内部に存在する、部分ペプチド鎖の除去を受け、その両端が連結されてなる短縮されたペプチド鎖からなるタンパク質であると想定する。

**【0108】**

その場合、上述する前還元処理済みの直鎖状ペプチド鎖とし、部位特異的なタンパク質分解処理を施して、複数のペプチド断片とした際、質量分析において、各ペプチド断片の「親イオン種」の分子量に関しては、かかる除去を受けている部分ペプチド鎖の両端の連結部を含むペプチド断片の「親イオン種」の分子量は、完全長アミノ酸配列に基づき推定される推定ペプチド断片の分子量のいずれとも相違したものとなる。当然、かかる除去を受けている部分ペプチド鎖内における、部位特異的なタンパク質分解処理によって、その断片化が引き起こされるペプチド断片に由来する「親イオン種」は、観測されることはない。

**【0109】**

仮に、前記参照基準データ・ベース中に含まれる公知タンパク質の一つに対する（推断）完全長アミノ酸配列が、解析対象のタンパク質の完全長アミノ酸配列と同一のアミノ酸配列を有するとすれば、上述する一次対比操作において、少なくとも、かかる公知タンパク質に由来するセット中の、推定ペプチド断片複数がある推定分子量 (Mref) と実質的に一致すると判定される、該解析対象タンパク質由来の実測ペプチド断片数 (Nex-id



）ならびに該公知タンパク質由来の推定ペプチド断片数（Nref-id）を決定すると、かかる除去を受けている部分ペプチド鎖の両端の連結部を含むペプチド断片を除き、一致すると判定されるため、原理的には、該解析対象タンパク質由来の実測ペプチド断片総数（Nex）から 1 を減じた数となる。

#### 【0 1 1 0】

該解析対象のタンパク質中の、除去を受けている部分ペプチド鎖の両端の連結部を含むペプチド断片の分子量と、偶々、同じ分子量を示す、推定ペプチド断片を有し、残る該解析対象タンパク質由来の実測ペプチド断片数（Nex-1）に関しても、それらの質量実測値（Mex）と一致する、推定ペプチド断片複数が有する推定分子量（Mref）を示す、別種の公知タンパク質の存在は、完全には、排除できないものの、そのような確率は相当に低い。

#### 【0 1 1 1】

従って、前記参照基準データ・ベース中に含まれる公知タンパク質の一つに対する（推断）完全長アミノ酸配列が、解析対象のタンパク質の完全長アミノ酸配列と同一のアミノ酸配列を有するとすれば、少なくとも、前記一次対比操作において、選別される、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群の中に、極めて高い確率で、かかる解析対象のタンパク質の完全長アミノ酸配列と同一のアミノ酸配列を有する、（推断）完全長アミノ酸配列を有する公知タンパク質が含まれることになる。その際、相当に高い確率で、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群は、この解析対象のタンパク質の完全長アミノ酸配列と同一のアミノ酸配列を有する、（推断）完全長アミノ酸配列を有する公知タンパク質のみとなる。換言するならば、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群に含まれる公知タンパク質種類が一つである場合、このデータ・ベース中より選別される該一種類の公知タンパク質を、該解析対象タンパク質に対する、単一の同定の候補と判定することができる。

#### 【0 1 1 2】

また、単一の同定の候補として選別された公知タンパク質に対する（推断）完全長アミノ酸配列が、解析対象のタンパク質の完全長アミノ酸配列と同一のアミノ酸配列を有する際には、一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される該公知タンパク質の配列情報を参照し、

一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量（Mref）と、一致すると判定されている、該解析対象タンパク質由来の実測ペプチド断片複数について、対応している該公知タンパク質に由来する推定ペプチド断片が占めるべき位置に配置した際、該解析対象タンパク質由来のペプチド断片が全て検出されている場合には、一致したと判定されていない推定ペプチド断片は、一連の部分領域を占め、この一連の未同定領域を除き、一致すると判定される前記実測ペプチド断片の一群が、該公知タンパク質の完全長アミノ酸配列中に含まれる、連続したアミノ酸配列を構成しているはずである。

#### 【0 1 1 3】

場合によっては、一致したと判定されていない推定ペプチド断片が占める、一連の未同定領域が N 末端から始まり、一致すると判定される前記実測ペプチド断片の一群は、かかる N 末側を除き、該公知タンパク質の完全長アミノ酸配列中の C 末端へと続く、連続したアミノ酸配列を構成している、あるいは、逆に、一致したと判定されていない推定ペプチド断片が占める、一連の未同定領域が C 末端に位置し、一致すると判定される前記実測ペプチド断片の一群は、かかる C 末側を除き、該公知タンパク質の完全長アミノ酸配列中の N 末端から続く、連続したアミノ酸配列を構成していることもある。この一致すると判定される前記実測ペプチド断片の一群が、連続したアミノ酸配列を構成している場合には、かかる一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

## 【0 1 1 4】

また、一致すると判定される前記実測ペプチド断片は、N末端側の一連の領域と、C末端側の一連の領域を占め、その間を、一致したと判定されていない推定ペプチド断片が占める、一連の未同定領域が埋めており、この合計3領域に、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質の（推断）完全長アミノ酸配列が分割されている場合にも、かかる一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

## 【0 1 1 5】

加えて、該解析対象タンパク質由来のペプチド断片が全て検出されている場合には、前記一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量（Mref）と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が唯一つ残っており、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分と連結される、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定である、一連の推定ペプチド断片群が占める部分について、翻訳後、その一連の未同定領域中に、プロテイン・スプライシング過程による部分切除が生じ、かかるタンパク質へと変換されると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、プロテイン・スプライシング過程に由来する一連の推定ペプチド断片複数の推定分子量（Mref）を算出し、

前記プロテイン・スプライシング過程に由来する一連の推定ペプチド断片に対する推定分子量（Mref）のうち、唯一残されている未同定の該解析対象タンパク質由来実測ペプチド断片の質量実測値（Mex）と一致するものの有無を判定する、二次対比操作を行うと、

前記プロテイン・スプライシング過程に由来する一連の推定ペプチド断片に対する推定分子量（Mref）のうち、唯一残されている未同定の該解析対象タンパク質由来実測ペプチド断片の質量実測値（Mex）と一致するものが一つ選別されるはずである。実際に、かかる二次対比操作において、一致するものの存在が検証された場合、上記の一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

## 【0 1 1 6】

具体的には、かかる除去を受けている部分ペプチド鎖の両端の連結部を含むペプチド断片は、一連の未同定領域中、そのN末側に位置する推定ペプチド断片のN末側部分アミノ酸配列と、C末側に位置する推定ペプチド断片のC末側部分アミノ酸配列とが連結されたものであり、この特徴に基づき、プロテイン・スプライシング過程に由来する一連の推定ペプチド断片複数の推定分子量（Mref）を容易に算出することが可能である。

## 【0 1 1 7】

偶然、該解析対象のタンパク質中の、プロテイン・スプライシング過程により除去を受けて、部分ペプチド鎖の両端が連結された部位と、部位特異的なタンパク質分解処理による切断が起こる部位とが一致することもある。その場合、結果的に、一次対比操作の際、未同定の該解析対象タンパク質由来実測ペプチド断片が残らないことになり、かかる場合も、当然、かかる一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

## 【0 1 1 8】

(F) alternative splicingに起因するスプライシング変異体型タンパク質の同定

図1に例示するように、解析対象のタンパク質は、ゲノム遺伝子上にコードされる、一連のエキソン複数のうち、alternative splicingに起因して、エキソン一つまたはそれ以



上を含む翻訳枠が欠落した mRNA に基づき翻訳される、完全長アミノ酸配列を有するペプチド鎖からなるスプライシング変異体型タンパク質であると想定する。

#### 【0119】

その場合、上述する前処理済みの直鎖状ペプチド鎖とし、部位特異的なタンパク質分解処理を施して、複数のペプチド断片とした際、質量分析において、各ペプチド断片の「親イオン種」の分子量に関しては、かかる欠落している、エキソン一つまたはそれ以上を含む翻訳枠内のアミノ酸配列部分から、部位特異的なタンパク質分解処理によって、その断片化が引き起こされるはずのペプチド断片は、元々無く、かかるペプチド断片に由来する「親イオン種」は、観測されることはない。また、alternative splicing に起因して、連結された二つのエキソンの接合部にコードされるアミノ酸残基を含むペプチド断片の「親イオン種」の分子量は、一般的に、この種の alternative splicing が生じていない際に得られる、(推断) 完全長アミノ酸配列に基づき推定される推定ペプチド断片の分子量のいずれとも相違したものとなる。

#### 【0120】

仮に、前記参照基準データ・ベース中に含まれる公知タンパク質の一つに対する(推断) 完全長アミノ酸配列が、解析対象のタンパク質をコードするゲノム遺伝子上にコードされる、この種の alternative splicing が生じていない際に得られる完全長アミノ酸配列と同一のアミノ酸配列を有するとすれば、上述する一次対比操作において、少なくとも、かかる公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と実質的に一致すると判定される、該解析対象タンパク質由来の実測ペプチド断片数 (Nex-id) ならびに該公知タンパク質由来の推定ペプチド断片数 (Nref-id) を決定すると、かかる alternative splicing に起因して、連結された二つのエキソンの接合部にコードされるアミノ酸残基を含むペプチド断片を除き、一致すると判定されるため、原理的には、該解析対象タンパク質由来の実測ペプチド断片総数 (Nex) から 1 を減じた数となる。

#### 【0121】

該解析対象のタンパク質中の、該 alternative splicing に起因して、連結された二つのエキソンの接合部にコードされるアミノ酸残基を含むペプチド断片の分子量と、偶々、同じ分子量を示す、推定ペプチド断片を有し、残る該解析対象タンパク質由来の実測ペプチド断片数 (Nex-1) に関しても、それらの質量実測値 (Mex) と一致する、推定ペプチド断片複数が有する推定分子量 (Mref) を示す、別種の公知タンパク質の存在は、完全には、排除できないものの、そのような確率は相当に低い。

#### 【0122】

従って、前記参照基準データ・ベース中に含まれる公知タンパク質の一つに対する(推断) 完全長アミノ酸配列が、解析対象のタンパク質をコードするゲノム遺伝子上にコードされる、この種の alternative splicing が生じていない際に得られる完全長アミノ酸配列と同一のアミノ酸配列を有するとすれば、少なくとも、前記一次対比操作において、選別される、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群の中に、極めて高い確率で、かかる解析対象のタンパク質の完全長アミノ酸配列と同一のアミノ酸配列を有する、(推断) 完全長アミノ酸配列を有する公知タンパク質が含まれることになる。その際、相当に高い確率で、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群は、この解析対象タンパク質をコードするゲノム遺伝子上にコードされる、この種の alternative splicing が生じていない際に得られる完全長アミノ酸配列と同一のアミノ酸配列を有する、(推断) 完全長アミノ酸配列を有する公知タンパク質のみとなる。換言するならば、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群に含まれる公知タンパク質種類が一つである場合、このデータ・ベース中より選別される該一種類の公知タンパク質を、該解析対象タンパク質に対する、単一の同定の候補と判定することができる。

#### 【0123】

また、単一の同定の候補として選別された公知タンパク質に対する(推断) 完全長アミ

ノ酸配列が、解析対象タンパク質をコードするゲノム遺伝子上にコードされる、この種の alternative splicing が生じていない際に得られる完全長アミノ酸配列と同一のアミノ酸配列を有する際には、一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される該公知タンパク質の配列情報を参照し、

一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致すると判定されている、該解析対象タンパク質由来の実測ペプチド断片複数について、対応している該公知タンパク質に由来する推定ペプチド断片が占めるべき位置に配置した際、該解析対象タンパク質由来のペプチド断片が全て検出されている場合には、一致したと判定されていない推定ペプチド断片は、一連の部分領域を占め、この一連の未同定領域を除き、一致すると判定される前記実測ペプチド断片の一群が、該公知タンパク質の完全長アミノ酸配列中に含まれる、連続したアミノ酸配列を構成しているはずである。

#### 【0124】

場合によっては、一致したと判定されていない推定ペプチド断片が占める、一連の未同定領域が N 末端から始まり、一致すると判定される前記実測ペプチド断片の一群は、かかる N 末側を除き、該公知タンパク質の完全長アミノ酸配列中の C 末端へと続く、連続したアミノ酸配列を構成している、あるいは、逆に、一致したと判定されていない推定ペプチド断片が占める、一連の未同定領域が C 末端に位置し、一致すると判定される前記実測ペプチド断片の一群は、かかる C 末側を除き、該公知タンパク質の完全長アミノ酸配列中の N 末端から続く、連続したアミノ酸配列を構成していることもある。この一致すると判定される前記実測ペプチド断片の一群が、連続したアミノ酸配列を構成している場合には、かかる一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0125】

また、一致すると判定される前記実測ペプチド断片は、N 末端側の一連の領域と、C 末側の一連の領域を占め、その間を、一致したと判定されていない推定ペプチド断片が占める、一連の未同定領域が埋めており、この合計 3 領域に、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質の (推断) 完全長アミノ酸配列が分割されている場合にも、かかる一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0126】

加えて、該解析対象タンパク質由来のペプチド断片が全て検出されている場合には、前記一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量 (Mref) と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が唯一つ残っており、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分と連結される、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定である、一連の推定ペプチド断片群が占める部分について、翻訳後、その一連の未同定領域中に含まれているアミノ酸配列部分をコードしている、一連のエキソン一つまたはそれ以上の翻訳枠が、alternative splicing 過程に起因して、含まれていない mRNA から翻訳されたスプライシング変異体型タンパク質と仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、スプライシング変異体型タンパク質に固有な一連の推定ペプチド断片複数の推定分子量 (Mref) を算出し、

前記スプライシング変異体型タンパク質に固有な一連の推定ペプチド断片に対する推定分子量 (Mref) のうち、唯一残されている未同定の該解析対象タンパク質由来実測ペプチド断片の質量実測値 (Mex) と一致するものの有無を判定する、二次対比操作を行うと



前記スプライシング変異体型タンパク質に固有な一連の推定ペプチド断片に対する推定分子量 (Mref) のうち、唯一残されている未同定の該解析対象タンパク質由来実測ペプチド断片の質量実測値 (Mex) と一致するものが一つ選別されるはずである。実際に、かかる二次対比操作において、一致するものの存在が検証された場合、上記の一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0127】

具体的には、かかるスプライシング変異体型タンパク質に固有なペプチド断片は、一連の未同定領域中、そのN末側に位置する推定ペプチド断片のN末側部分アミノ酸配列と、C末側に位置する推定ペプチド断片のC末側部分アミノ酸配列とが連結されたものであり、その連結部は、alternative splicing過程に起因して、連結される二つのエキソンの接合部にコードされるアミノ酸残基に相当するという特徴に基づき、スプライシング変異体型タンパク質に固有な一連の推定ペプチド断片複数の推定分子量 (Mref) を容易に算出することが可能である。

#### 【0128】

偶然、該解析対象のタンパク質中の、alternative splicing過程に起因して、連結される二つのエキソンの接合部にコードされるアミノ酸残基部位と、部位特異的なタンパク質分解処理による切断が起こる部位とが一致することもある。その場合、結果的に、一次対比操作の際、未同定の該解析対象タンパク質由来実測ペプチド断片が残らないことになり、かかる場合も、当然、かかる一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0129】

(G) 参照するデータ・ベースにおける、(推断) 完全長アミノ酸配列に誤謬を有する際における、タンパク質の同定

解析対象のタンパク質は、ゲノム遺伝子上にコードされる完全長アミノ酸配列を有するペプチド鎖からなるタンパク質であり、参照するデータ・ベース上に、かかる解析対象のタンパク質のゲノム遺伝子塩基配列は、公知タンパク質として、収録されているものの、かかるゲノム遺伝子によりコードされている、(推断) 完全長アミノ酸配列は、誤謬を有していると想定する。

#### 【0130】

例えば、対応するmRNA、またはそのcDNAの塩基配列解析はなされていないが、ゲノム遺伝子上に見出される、オープン・リーディング領域複数を仮想的に連結して、全体の翻訳枠を構築し、この仮想的なコード塩基配列に基づき、(推断) 完全長アミノ酸配列を仮に決定した配列情報も、参照するデータ・ベース上に収録されていることも少なくない。このような仮想的なコード塩基配列の構築過程では、連結すべきオープン・リーディング領域の選択に、複数の可能性があり、それぞれ、一連のコード塩基配列を与えている場合であっても、その全てを収録していないことも、少なくない。そのため、参照するデータ・ベース上に収録されている、仮想的なコード塩基配列自体は、合理的に構築されているが、現実には、収録されていない別の仮想的なコード塩基配列によって、ペプチド鎖への翻訳が起こっているといった、結果的に、同定上の誤謬がデータ・ベース中に存在することも想定できる。すなわち、図1の(1)に示すように、エキソン領域の想定複数のうち、収録されているものと、本来のものとは一部相違しているような場合である。

#### 【0131】

この種のエキソン領域の特定に誤謬を有する結果、参照するデータ・ベースにおける、(推断) 完全長アミノ酸配列に誤謬を有する際においては、参照基準データ・ベース中に含まれる該公知タンパク質に対する、エキソン領域の特定に部分的な誤謬を含む、仮想的な(推断) 完全長アミノ酸配列は、対応する一連のエキソン領域によりコードされるアミノ酸配列部分が、現実のものとは相違したものである。この誤ったアミノ酸配列部分を内

在する、仮想的な（推断）完全長アミノ酸配列に基づき推定される、該公知タンパク質に対する推定ペプチド断片複数が有する推定分子量（Mref）と、該解析対象タンパク質由来のペプチド断片の質量実測値（Mex）とを対比する、一次対比操作において、かかる誤ったアミノ酸配列部分に相当する、一連の推定ペプチド断片複数は、一致する該解析対象タンパク質由来の実測ペプチド断片は、当然に存在しないものとなる。一方、その誤ったアミノ酸配列部分を除く領域では、該公知タンパク質に対する推定ペプチド断片複数が有する推定分子量（Mref）と、該解析対象タンパク質由来のペプチド断片の質量実測値（Mex）とは、完全に一致する。

#### 【0132】

従って、前記参照基準データ・ベース中に含まれる、本来は、解析対象タンパク質と完全に一致すると判定されるべき、該公知タンパク質に対する、推定ペプチド断片複数が有する推定分子量（Mref）と実質的に一致すると判定される、該解析対象タンパク質由来の実測ペプチド断片数（Nex-id）ならびに該公知タンパク質由来の推定ペプチド断片数（Nref-id）を決定すると、該解析対象タンパク質由来の実測ペプチド断片総数（Nex）から、前記の誤ったアミノ酸配列部分に相当する、一連のペプチド断片数を減じた数となる。少なくとも、前記一次対比操作において、選別される、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群の中に、十分に高い確率で、かかる部分的にアミノ酸配列の誤りを持つ該公知タンパク質が含まれることになる。その際、相当に高い確率で、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群は、この部分的にアミノ酸配列の誤りを持つ公知タンパク質のみとなる。換言するならば、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群に含まれる公知タンパク質種類が一つである場合、このデータ・ベース中より選別される該一種類の公知タンパク質を、該解析対象タンパク質に対する、単一の同定の候補と判定することができる。

#### 【0133】

また、単一の同定の候補として選別された、該公知タンパク質に対する、部分的にアミノ酸配列の特定に誤りを有している（推断）完全長アミノ酸配列情報を参照し、一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量（Mref）と、一致すると判定されている、該解析対象タンパク質由来の実測ペプチド断片複数について、対応している該公知タンパク質に由来する推定ペプチド断片が占めるべき位置に配置した際、一致したと判定されていない推定ペプチド断片は、一連の部分領域を占め、この一連の未同定領域を除き、一致すると判定される前記実測ペプチド断片の一群が、該公知タンパク質の、部分的にアミノ酸配列の特定に誤りを有している（推断）完全長アミノ酸配列中においても、連続したアミノ酸配列を構成しているはずである。

#### 【0134】

通常は、一致すると判定される前記実測ペプチド断片は、N末端側の一連の領域と、C末端側の一連の領域を占め、その間を、一致したと判定されていない推定ペプチド断片が占める、一連の未同定領域が埋めており、この合計3領域に、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質の仮想的な（推断）完全長アミノ酸配列が分割されている場合にも、かかる一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0135】

前記（A）の事例と同様に、解析対象のタンパク質由来のペプチド断片各質量実測値（Mex）と、公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量（Mref）とが実質的に一致すると判定される場合、MS/MS分析法によって得られる、解析対象のタンパク質由来の各ペプチド断片に対応する一価の「親の陽イオン種」、または一価の「親の陰イオン種」について、その断片化により生成する各種の娘イオン種の分子量の測定結果と、ペプチド断片の分子量が一致すると判定される、公知タンパク



質に由来する推定ペプチド断片が有するアミノ酸配列から、MS/MS分析において、その断片化により生成すると想定される各種の娘イオン種の分子量推定値との間において、その対応を確認することにより、より確度の高い判定が可能となる。さらには、MS/MS分析法によって得られる、解析対象のタンパク質由来の各ペプチド断片に対応する一価の「親の陽イオン種」、または一価の「親の陰イオン種」について、その断片化により生成する各種の娘イオン種の分子量の測定結果に代えて、あるいは、さらに付加的に、前記の「ペプチドのC末端アミノ酸配列解析方法」の手法を利用して入手される、解析対象のタンパク質由来の各ペプチド断片におけるC末端側のアミノ酸配列情報をも、二次的な質量分析結果として、利用することで、公知タンパク質に由来する推定ペプチド断片が有するアミノ酸配列との部分的な一致も確認でき、より確度の高い単一の同定の候補を、選別することができる。

#### 【0136】

(H) 「一塩基多型」に起因する、アミノ酸置換が生じている変異体タンパク質の同定

解析対象のタンパク質は、ゲノム遺伝子上にコードされる完全長アミノ酸配列を有するペプチド鎖からなるタンパク質であり、かかる完全長アミノ酸配列中には、「一塩基多型」に起因する、アミノ酸置換が生じている変異体タンパク質であり、一方、参照するデータ・ベース上には、解析対象タンパク質のゲノム遺伝子において、前記「一塩基多型」に起因して、別のアミノ酸がコードされているものが、公知タンパク質として、収録されていると想定する。

#### 【0137】

その場合、上述する前処理済みの直鎖状ペプチド鎖とし、部位特異的なタンパク質分解処理を施して、複数のペプチド断片とした際、質量分析において、各ペプチド断片の「親イオン種」の分子量に関しては、解析対象タンパク質に由来するペプチド断片の質量実測値(Mex)と、公知タンパク質の(推断)完全長アミノ酸配列に基づき推定される推定ペプチド断片の分子量とは、前記「一塩基多型」に起因するアミノ酸の相違が生じているペプチド断片のみで、相違したものとなる。

#### 【0138】

前記参照基準データ・ベース中に含まれる、前記「一塩基多型」変異体複数種の一つである公知タンパク質と対比した際、上述する一次対比操作において、少なくとも、かかる公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量(Mref)と実質的に一致すると判定される、該解析対象タンパク質由来の実測ペプチド断片数(Nex-id)ならびに該公知タンパク質由来の推定ペプチド断片数(Nref-id)を決定すると、原理的には、該解析対象のタンパク質に由来し、前記「一塩基多型」のアミノ酸変異を含むペプチド断片数(Nex-snp)を、該解析対象タンパク質由来の実測ペプチド断片総数(Nex)から減じた数となる。

#### 【0139】

該解析対象のタンパク質に由来し、「一塩基多型」のアミノ酸変異を含むペプチド断片の分子量と、偶々、同じ分子量を示す、別種のアミノ酸配列を有するペプチド断片の存在は、完全には、排除できないものの、そのような確率は相当に低い。

#### 【0140】

また、該解析対象タンパク質に由来し、「一塩基多型」のアミノ酸変異を含むペプチド断片の分子量と、偶々、同じ分子量を示す、推定ペプチド断片を有し、残る該解析対象タンパク質由来の実測ペプチド断片数(Nex-Nex-snp)に関しても、それらの質量実測値(Mex)と一致する、推定ペプチド断片複数が有する推定分子量(Mref)を示す、別種の公知タンパク質の存在は、完全には、排除できないものの、そのような確率は極めて低い。

#### 【0141】

従って、前記参照基準データ・ベース中に含まれる、前記「一塩基多型」変異体複数種の一つである公知タンパク質は、少なくとも、前記一次対比操作において、選別される、

該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群の中に、極めて高い確率で含まれることになる。その際、相当に高い確率で、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群は、この解析対象タンパク質と対応するゲノム遺伝子を共通とする、「一塩基多型」変異体複数種の一つである公知タンパク質のみとなる。換言するならば、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質群に含まれる公知タンパク質種類が一つである場合、このデータ・ベース中より選別される該一種類の公知タンパク質を、該解析対象タンパク質に対する、単一の同定の候補と判定することができる。

#### 【0 1 4 2】

また、単一の同定の候補として選別された、「一塩基多型」変異体複数種の一つである公知タンパク質に対する（推断）完全長アミノ酸配列を参照し、

一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量（Mref）と、一致すると判定されている、該解析対象タンパク質由来の実測ペプチド断片複数について、対応している該公知タンパク質に由来する推定ペプチド断片が占めるべき位置に配置した際、一致したと判定されていない推定ペプチド断片は、「一塩基多型」に起因して、アミノ酸が相違している部分領域を反映しているはずである。

#### 【0 1 4 3】

仮に、「一塩基多型」に起因してアミノ酸が変換され、部位特異的なタンパク質分解処理の切断部位が消失すると、かかる切断部位によって、分離されていた二つの断片が一体化したペプチド断片が得られる。また、「一塩基多型」に起因してアミノ酸が変換され、部位特異的なタンパク質分解処理の切断部位が生じていると、かかる切断部位によって、一つのペプチド断片から、派生する二つのペプチド断片が得られる。

#### 【0 1 4 4】

部位特異的なタンパク質分解処理の切断部位の消失、生成を伴わない、「一塩基多型」に起因したアミノ酸変換では、そのペプチド断片の分子量は、アミノ酸種の差違に相当する変化を引き起こす。

#### 【0 1 4 5】

（H-1） 「一塩基多型」に起因してアミノ酸が変換され、部位特異的なタンパク質分解処理の切断部位が消失する場合

図5に例示するように、かかる切断部位によって、分離されていた二つの断片が一体化したペプチド断片となる結果、前記一次対比操作において、一致したと判定されていない、公知タンパク質由来の推定ペプチド断片複数のうち、少なくとも隣接する二つの推定ペプチド断片が見出される。その二つの推定ペプチド断片が連結された状態において推定される分子量（Mref-ad）と、類似する質量実測値（Mex）を示す、該解析対象タンパク質に由来する未同定のペプチド断片一つが存在するはずである。また、変異するアミノ酸残基自体（Xref-snp）は、前記隣接する二つの推定ペプチド断片のアミノ酸配列から判明している。推定される分子量（Mref-ad）と質量実測値（Mex）との差違 $\Delta Mad$ （ $= Mref-ad - Mex$ ）と、変異するアミノ酸残基（Xref-snp）から、変異されたアミノ酸残基（Xex-snp）を推断することが可能である。さらに、公知タンパク質において、報告されているゲノム遺伝子塩基配列において、該変異するアミノ酸残基（Xref-snp）をコードするコドン配列を参照し、実際に、「一塩基多型」に起因して、変異されたアミノ酸残基（Xex-snp）をコードする、コドン配列への変換が可能であることを確認する。

#### 【0 1 4 6】

（H-2） 「一塩基多型」に起因してアミノ酸が変換され、部位特異的なタンパク質分解処理の切断部位が生成する場合

図4に例示するように、かかる切断部位によって、一つのペプチド断片から、派生する二つのペプチド断片が得られているはずであり、前記一次対比操作において、一致したと判定されていない、公知タンパク質由来の推定ペプチド断片複数のうち、少なくとも、その喪失すべき推定ペプチド断片の推定分子量（Mref）と類似する質量実測値（Mex）を



示す、該解析対象タンパク質に由来する未同定のペプチド断片は、全く存在していないはずである。すなわち、部位特異的なタンパク質分解処理の切断部位が生成されたにも係わらず、実際には、切断が生じていないような、該解析対象タンパク質に由来する未同定のペプチド断片は、存在していないはずである。

#### 【0147】

一方、その喪失すべき推定ペプチド断片中に、切断部位が生成された結果、派生する二つのペプチド断片の有する分子量 ( $M_{ex-fra1}$ 、 $M_{ex-fra2}$ ) は、当然に、喪失すべき推定ペプチド断片の推定分子量 ( $M_{ref}$ ) より小さな値となり、また、この派生する二つのペプチド断片を連結したペプチド断片が示すはずの分子量 ( $M_{ex-fra1}+fra2$ ) は、 $M_{ex-fra1}+M_{ex-fra2}-18$ 、すなわち、前記派生する二つのペプチド断片の有する分子量の和より、アミノ結合形成による水一分子分の式量 (18) を減じたものとなる。勿論、この  $M_{ex-fra1}+M_{ex-fra2}-18$  は、かかる喪失すべき推定ペプチド断片の推定分子量 ( $M_{ref}$ ) と類似する値である。

#### 【0148】

喪失すべき推定ペプチド断片の推定分子量 ( $M_{ref}$ ) より小さな値の、質量実測値 ( $M_{ex}$ ) を示す、該解析対象タンパク質に由来する未同定のペプチド断片複数より、前記の要件を満足する、二つのペプチド断片を選別することができるはずである。選別された二つのペプチド断片の有する質量実測値 ( $M_{ex}$ ) に基づき、派生する二つのペプチド断片を連結したペプチド断片が示すはずの分子量 ( $M_{ex-fra1}+fra2$ ) = ( $M_{ex-fra1}+M_{ex-fra2}-18$ ) に相当する値を算出し、それと喪失すべき推定ペプチド断片の推定分子量 ( $M_{ref}$ ) との差違  $\Delta M_{div}$  (=  $M_{ref}-M_{ex-fra1}+fra2$ ) を算出する。

#### 【0149】

一方、変異するアミノ酸残基 ( $X_{ref-snp}$ ) は判明していないが、変異されたアミノ酸残基 ( $X_{ex-snp}$ ) は、部位特異的なタンパク質分解処理の切断部位を与えるもので、判明している。従って、差違  $\Delta M_{div}$  (=  $M_{ref}-M_{ex-fra1}+fra2$ ) と、変異されたアミノ酸残基 ( $X_{ex-snp}$ ) とから、変異するアミノ酸残基 ( $X_{ref-snp}$ ) を推断することが可能である。実際に、公知タンパク質由来の喪失すべき推定ペプチド断片のアミノ酸配列中に、かかる推断された変異するアミノ酸残基 ( $X_{ref-snp}$ ) が存在しており、これを変異されたアミノ酸残基 ( $X_{ex-snp}$ ) へと変換した際、派生する二つのペプチド断片の推定分子量と、該解析対象タンパク質に由来する未同定のペプチド断片群から選別された二つのペプチド断片の分子量 ( $M_{ex-fra1}$ 、 $M_{ex-fra2}$ ) とが対応していることを確認する。さらに、公知タンパク質において、報告されているゲノム遺伝子塩基配列において、該変異するアミノ酸残基 ( $X_{ref-snp}$ ) をコードするコドン配列を参照し、実際に、「一塩基多型」に起因して、変異されたアミノ酸残基 ( $X_{ex-snp}$ ) をコードする、コドン配列への変換が可能であることを確認する。

#### 【0150】

(H-3) 部位特異的なタンパク質分解処理の切断部位の消失、生成を伴わない、「一塩基多型」に起因したアミノ酸変換のみが起きている場合

かかる部位特異的なタンパク質分解処理の切断部位の消失、生成を伴わない、「一塩基多型」に起因したアミノ酸変換では、そのペプチド断片の分子量は、アミノ酸種の差違に相当する変化を引き起こす。

#### 【0151】

前記一次対比操作において、一致したと判定されていない、公知タンパク質由来の推定ペプチド断片複数のうち、少なくとも、推定ペプチド断片の推定分子量 ( $M_{ref}$ ) の一つと、類似する質量実測値 ( $M_{ex}$ ) を示す、該解析対象タンパク質に由来する未同定のペプチド断片一つが存在するはずである。

#### 【0152】

具体的には、一つのアミノ酸変換に起因する分子量変化  $\Delta M_{xy}$  は、トリプトファン (Trp) とグリシン (Gly) の式量差 129 を超えることはない。また、変異するアミノ酸残基 ( $X_{ref-snp}$ )、変異されたアミノ酸残基 ( $X_{ex-snp}$ ) のいずれも、部位特異的なタンパ

ク質分解処理の切断部位を与えるアミノ酸残基とは相違するはずである。

#### 【0153】

一次対比操作において、未同定となっている、公知タンパク質由来の推定ペプチド断片について、分子量差が129の範囲内に、該解析対象タンパク質に由来する未同定のペプチド断片が存在しているか否かを判別する。存在していると判別された該解析対象タンパク質に由来する未同定のペプチド断片に関して、両者間の分子量差 $\Delta M_{ref-ex}$ を算出する。

#### 【0154】

公知タンパク質由来の推定ペプチド断片の有する（推断）アミノ酸配列は、判明しているので、そのアミノ酸配列中に含まれるアミノ酸を変換した際、分子量差 $\Delta M_{ref-ex}$ を与えるアミノ酸変換の有無を判定する。仮に、そのようなアミノ酸変換が複数存在する際には、公知タンパク質において、報告されているゲノム遺伝子塩基配列において、該変異するアミノ酸残基（Xref-snp）をコードするコドン配列を参照し、単一の「一塩基多型」部位にみに起因して、変異されたアミノ酸残基（Xex-snp）をコードする、コドン配列への変換が可能であるか否かを確認する。すなわち、アミノ酸変換が、一つの塩基を変化するのみで達成されるもの、例えば、GTGでコードされるValからCTGでコードされるLeuへの変換などが、より確度が高く、逆に、GGGでコードされるGlyからTTTでコードされるPheへの変換は、相当に確度は低いと判断され、より確度の高い変換を、第一の候補に選択する。なお、実際に、解析対象タンパク質において、どのようなコドンでコードされているかは、かかる解析対象タンパク質の具体的な起源である個体における、mRNAに基づき、そのコード塩基配列の確認を行うことが必要である。

#### 【0155】

以上の手順に従って、「一塩基多型」に起因する、アミノ酸置換が生じている変異体タンパク質と想定される際、

一次対比操作において、該同定の候補と判定される公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量（Mref）と、一致するとは判定されていない、未同定の該解析対象タンパク質由来実測ペプチド断片が残っている場合、該未同定の該解析対象タンパク質由来実測ペプチド断片に関して、

該公知タンパク質の完全長アミノ酸配列中に含まれる、前記連続したアミノ酸配列部分中に存在する、該同定の候補と判定される公知タンパク質に由来し、対応する実測ペプチド断片が未同定である、内部未同定領域の推定ペプチド断片の群の部分に対して、それをコードするゲノム遺伝子部分において、該ゲノム遺伝子部分に含まれるエキソン中に、一塩基多型に起因する、翻訳されるアミノ酸置換が一つ生じていると仮定した際に、想定されるアミノ酸配列に基づき、前記保護基の導入処理と、部位特異的なタンパク質分解処理とを施した際に生成すると予測される、一塩基多型のアミノ酸置換に由来する推定ペプチド断片複数の推定分子量（Mref）を算出し、

前記一塩基多型のアミノ酸置換に由来する推定ペプチド断片に対する推定分子量（Mref）と一致する、質量実測値（Mex）を有する未同定の該解析対象タンパク質由来実測ペプチド断片の有無を判定する、二次対比操作を行い、

前記一塩基多型のアミノ酸置換に由来する推定ペプチド断片に対する推定分子量（Mref）と一致する、質量実測値（Mex）を有する未同定の該解析対象タンパク質由来実測ペプチド断片が少なくとも一つ選別される場合、

一次対比操作の結果に基づき選別される、該解析対象タンパク質に対する、単一の同定の候補と判定される公知タンパク質は、より高い確度で単一の同定の候補と判定することが可能である。

#### 【0156】

前記（A）の事例と同様に、解析対象のタンパク質由来のペプチド断片各質量実測値（Mex）と、公知タンパク質に由来するセット中の、推定ペプチド断片複数が有する推定分子量（Mref）とが実質的に一致すると判定される場合、MS/MS分析法によって得られる、解析対象のタンパク質由来の各ペプチド断片に対応する一価の「親の陽イオン種」



、または一価の「親の陰イオン種」について、その断片化により生成する各種の娘イオン種の分子量の測定結果と、ペプチド断片の分子量が一致すると判定される、公知タンパク質に由来する推定ペプチド断片が有するアミノ酸配列から、MS/MS分析において、その断片化により生成すると想定される各種の娘イオン種の分子量推定値との間において、その対応を確認することにより、より確度の高い判定が可能となる。さらには、MS/MS分析法によって得られる、解析対象のタンパク質由来の各ペプチド断片に対応する一価の「親の陽イオン種」、または一価の「親の陰イオン種」について、その断片化により生成する各種の娘イオン種の分子量の測定結果に代えて、あるいは、さらに付加的に、前記の「ペプチドのC末端アミノ酸配列解析方法」の手法を利用して入手される、解析対象のタンパク質由来の各ペプチド断片におけるC末端側のアミノ酸配列情報をも、二次的な質量分析結果として、利用することで、公知タンパク質に由来する推定ペプチド断片が有するアミノ酸配列との部分的な一致も確認でき、より確度の高い単一の同定の候補を、選別することができる。

#### 【0157】

例えば、ヒトにおける、各アミノ酸をコードするコドンと、その使用頻度を表1に示す。「一塩基多型」に起因する、アミノ酸置換は、各個体において、そのゲノム遺伝子上において、特定部位に存在する一つの塩基が、複数の塩基種を採ることで、かかる塩基を含むコドンによりコードされるアミノ酸種も変化することに起因している。実際に、その種の「一塩基多型」による塩基配列の変異が、データ・ベースに副次的な情報として、収録されているものもあるが、本発明では、仮にそのような副次的な情報が収録されていない場合においても、可能性のある「一塩基多型」に起因する、アミノ酸置換を以下のような手順に従って推定することで、二次対比操作に利用する、仮想的な変異を有するペプチド断片のアミノ酸配列と、その推定分量を算出することもできる。

#### 【0158】

なお、「一塩基多型」に起因する、アミノ酸置換を考慮する際、各コドンのうち、一塩基の置換によって引き起こされる、コードされるアミノ酸変化は、下記表2～表13に列記するものである。この一塩基置換により引き起こされるアミノ酸置換を、表14に纏めて示す。その他、各コドンに含まれる塩基の二つ、三つが、置換されることで、コードされるアミノ酸変化が起こる可能性も排除できず、それらを含め、各アミノ酸相互間の変異を引き起こすに必要な、最小の塩基変異数を、表15に纏めて示す。

#### 【0159】

「一塩基多型」に起因する、アミノ酸置換が生じた際、そのアミノ酸間の式量差に相当する分子量の変化が観測されるはずである。各分子量変化量を与える、アミノ酸置換を纏めると、表16に示すようになる。その内、下線を付すアミノ酸置換は、一塩基置換により引き起こされるアミノ酸置換であり、「一塩基多型」に起因するアミノ酸置換として、より蓋然性の高い候補と考えられる。

#### 【0160】

従って、一次対比操作において、未同定とされた公知タンパク質由来の推定ペプチド断片に対して、そのアミノ酸配列に基づき、「一塩基多型」に起因するアミノ酸置換が生じた際に予測される推定分子量の算出は、表16に示す、アミノ酸置換と分子量変化量との関係を参照し、そのアミノ酸配列に含まれるアミノ酸種に基づき行い、一群の推定分子量とそれを与える一アミノ酸の変異を有するアミノ酸配列を決定する。なお、データ・ベースに収録される、該公知タンパク質のゲノム遺伝子において、かかるアミノ酸をコードしているコドンを確認し、一塩基置換により引き起こされるアミノ酸置換を有するもののみを、より優先度の高い、「一塩基多型」に起因するアミノ酸置換を有する推定ペプチド断片の群として、二次対比操作において利用することも可能である。

#### 【0161】

加えて、ゲノム遺伝子に存在する「一塩基多型」型の塩基配列変異を引き起こす要因、機構に関しては、現段階では、研究が進められている状態ではある。すなわち、ヒト、哺乳動物などの、有性生殖により、ゲノムの遺伝子情報が受け継がれる生物において、個々

の個体間において、「一塩基多型」型の塩基配列変異の特定例は、少なくないものの、その個々の「一塩基多型」型の塩基配列変異が導入される誘因、具体的な機構の解明に関しては、更なる研究を待たなければならない。しかしながら、一般的に、ゲノム遺伝子中の塩基配列変異は、ゲノム遺伝子の複製、あるいは、遺伝子損傷の修復過程において、本来の塩基と異なる塩基への変換が生じることに由来すると想定されている。

#### 【0162】

人為的に誘導される誘発突然変異の研究において、ゲノム遺伝子損傷に対する修復誤り、ゲノム遺伝子の複製時の、鋳型の一本鎖DNA上の塩基に軽微な損傷があることによる対合誤り、あるいは、複製自体の誤りを増加させた際、見出される塩基配列変異の類別化の研究成果から、前記の機構に由来する点突然変異、すなわち、塩基対置換における発生頻度に関する統計的な規則性（経験則）が示されている。すなわち、表現型自体にも変化が見いだされ、形質的变化を示している突然変異体においては、プリン塩基相互（ $A \leftrightarrow G$ ）、あるいは、ピリミジン塩基相互（ $T \leftrightarrow C$ ）の置換である transition は、プリン塩基（A, G）とピリミジン塩基（T, C）との間における置換である transversion より、大幅に高い頻度で見出されている。加えて、transition 型塩基対置換の間、あるいは、transversion 型塩基対置換の間における詳細な頻度比較を行うと、transition 型塩基対置換の間、あるいは、transversion 型塩基対置換の間においても、統計的に有意な差が存在している。これら見出された頻度の傾向を纏めると下記する順序となる。

#### 【0163】

$\text{transition } (T \leftrightarrow C, A \leftrightarrow G) > \text{transversion } (A \leftrightarrow C, T \leftrightarrow G, G \leftrightarrow C, A \leftrightarrow T)$

細かく分類すると、ゲノム遺伝子中のコード鎖における塩基配列においては、下記する順序となる。

#### 【0164】

$T \leftrightarrow C > A \leftrightarrow G > [A \leftrightarrow C, T \leftrightarrow G] > [G \leftrightarrow C, A \leftrightarrow T]$

一方、部位特異的なタンパク質分解処理の切断部位の消失、生成を伴わない、「一塩基多型」に起因したアミノ酸変換、例えば、部位特異的なタンパク質分解処理にトリプシンを利用する際には、リシン、アルギニン残基から他のアミノ酸残基への変異、あるいは、他のアミノ酸残基から他のアミノ酸残基への変異を除外し、各コドンのうち、一塩基の置換によって引き起こされる、コードされるアミノ酸変化が、同じ質量変化を引き起こす組み合わせが複数あるものを纏めると、表17に示すものが挙げられる。

#### 【0165】

表17に示す、これらのアミノ酸変換を引き起こすコドン変化を検討すると、下記のように、transition 型塩基対置換と transversion 型塩基対置換とに分類される。

#### 【0166】

・  $d = \pm 1$

$N \leftrightarrow D$ ;  $\underline{A}AT \leftrightarrow \underline{G}AT$ ,  $\underline{A}AC \leftrightarrow \underline{G}AC$ : ( $A \leftrightarrow G$ ) transition 型  
 $I \leftrightarrow N$ ;  $\underline{A}TT \leftrightarrow \underline{A}AT$ ,  $\underline{A}TC \leftrightarrow \underline{A}AC$ : ( $T \leftrightarrow A$ ) transversion 型  
 $Q \leftrightarrow E$ ;  $\underline{C}AA \leftrightarrow \underline{G}AA$ ,  $\underline{C}AG \leftrightarrow \underline{G}AG$ : ( $C \leftrightarrow G$ ) transversion 型

・  $d = \pm 16$

$P \leftrightarrow L$ ;  $\underline{C}CT \leftrightarrow \underline{C}TT$ ,  $\underline{C}CC \leftrightarrow \underline{C}TC$   
 $\underline{C}CA \leftrightarrow \underline{C}TA$ ,  $\underline{C}CG \leftrightarrow \underline{C}TG$ : ( $C \leftrightarrow T$ ) transition 型  
 $A \leftrightarrow S$ ;  $\underline{G}CT \leftrightarrow \underline{T}CT$ ,  $\underline{G}CC \leftrightarrow \underline{T}CC$   
 $\underline{G}CA \leftrightarrow \underline{T}CA$ ,  $\underline{G}CG \leftrightarrow \underline{T}CG$ : ( $G \leftrightarrow T$ ) transversion 型  
 $S \leftrightarrow C$ ;  $\underline{T}CT \leftrightarrow \underline{T}GT$ ,  $\underline{T}CC \leftrightarrow \underline{T}GC$ : ( $C \leftrightarrow G$ ) transversion 型  
 $\underline{A}GT \leftrightarrow \underline{T}GT$ ,  $\underline{A}GC \leftrightarrow \underline{T}GC$ : ( $A \leftrightarrow T$ ) transversion 型  
 $V \leftrightarrow D$ ;  $\underline{G}TT \leftrightarrow \underline{G}AT$ ,  $\underline{G}TC \leftrightarrow \underline{G}AC$ : ( $T \leftrightarrow A$ ) transversion 型  
 $F \leftrightarrow Y$ ;  $\underline{T}TT \leftrightarrow \underline{T}AT$ ,  $\underline{T}TC \leftrightarrow \underline{T}AC$ : ( $T \leftrightarrow A$ ) transversion 型



- ・  $d = \pm 26$   
 $S \Leftrightarrow L$ ;  $TCA \Leftrightarrow TTA$ ,  $TCG \Leftrightarrow TTG$ : (C $\Leftrightarrow$ T) transition型  
 $H \Leftrightarrow Y$ ;  $CAT \Leftrightarrow TAT$ ,  $CAC \Leftrightarrow TAC$ : (C $\Leftrightarrow$ T) transition型  
 $S \Leftrightarrow I$ ;  $AGT \Leftrightarrow ATT$ ,  $AGC \Leftrightarrow ATC$ : (G $\Leftrightarrow$ T) transversion型  
 $A \Leftrightarrow S$ ;  $GCT \Leftrightarrow CCT$ ,  $GCC \Leftrightarrow CCC$   
 $GCA \Leftrightarrow CCA$ ,  $GCG \Leftrightarrow CCG$ : (G $\Leftrightarrow$ C) transversion型
- ・  $d = \pm 30$   
 $T \Leftrightarrow M$ ;  $ACG \Leftrightarrow ATG$ : (C $\Leftrightarrow$ T) transition型  
 $G \Leftrightarrow S$ ;  $GGT \Leftrightarrow AGT$ ,  $GGC \Leftrightarrow AGC$ : (G $\Leftrightarrow$ A) transition型  
 $A \Leftrightarrow T$ ;  $GCT \Leftrightarrow ACT$ ,  $GCC \Leftrightarrow ACC$   
 $GCA \Leftrightarrow ACA$ ,  $GCG \Leftrightarrow ACG$ : (G $\Leftrightarrow$ A) transition型  
 $V \Leftrightarrow E$ ;  $GTA \Leftrightarrow GAA$ ,  $GTG \Leftrightarrow GAG$ : (T $\Leftrightarrow$ A) transversion型
- ・  $d = \pm 34$   
 $L \Leftrightarrow F$ ;  $CTT \Leftrightarrow TTT$ ,  $CTC \Leftrightarrow TTC$ : (C $\Leftrightarrow$ T) transition型  
 $I \Leftrightarrow F$ ;  $ATT \Leftrightarrow TTT$ ,  $ATC \Leftrightarrow TTC$ : (A $\Leftrightarrow$ T) transversion型
- ・  $d = \pm 44$   
 $C \Leftrightarrow F$ ;  $TGT \Leftrightarrow TTT$ ,  $TGC \Leftrightarrow TTC$ : (G $\Leftrightarrow$ T) transversion型  
 $A \Leftrightarrow D$ ;  $GCT \Leftrightarrow GAT$ ,  $GCC \Leftrightarrow GAC$ : (C $\Leftrightarrow$ A) transversion型
- ・  $d = \pm 48$   
 $V \Leftrightarrow F$ ;  $GTT \Leftrightarrow TTT$ ,  $GTC \Leftrightarrow TTC$   
 $GTA \Leftrightarrow TTA$ ,  $GTG \Leftrightarrow TTG$ : (G $\Leftrightarrow$ T) transversion型  
 $D \Leftrightarrow Y$ ;  $GAT \Leftrightarrow TAT$ ,  $GAC \Leftrightarrow TAC$ : (G $\Leftrightarrow$ T) transversion型
- ・  $d = \pm 58$   
 $G \Leftrightarrow D$ ;  $GGT \Leftrightarrow GAT$ ,  $GGC \Leftrightarrow GAC$ : (G $\Leftrightarrow$ A) transition型  
 $A \Leftrightarrow E$ ;  $GCA \Leftrightarrow GAA$ ,  $GCG \Leftrightarrow GAG$ : (C $\Leftrightarrow$ A) transversion型
- ・  $d = \pm 60$   
 $S \Leftrightarrow F$ ;  $TCT \Leftrightarrow TTT$ ,  $TCC \Leftrightarrow TTC$ : (C $\Leftrightarrow$ T) transition型  
 $C \Leftrightarrow Y$ ;  $TGT \Leftrightarrow TAT$ ,  $TGC \Leftrightarrow TAC$ : (G $\Leftrightarrow$ A) transition型

以上に示す、各アミノ酸変換を引き起こすコドン変化の出現頻度に関して、上述する点突然変異における統計的な頻度傾向に従うと仮定すると、表18に示す序列付けが可能である。従って、一塩基置換により引き起こされるアミノ酸置換を有するもののみを、より優先度の高い、「一塩基多型」に起因するアミノ酸置換を有する推定ペプチド断片の群として、二次対比操作において利用した際、場合によっては、表17に示す同一の質量変化を与える、複数の推定ペプチド断片が含まれることもあるが、その複数の推定ペプチド断片間における、より蓋然性の高い候補の選別に際して、表18に示す序列付けをも参考にすることができる。

【0167】

【表 1】

## コドン使用頻度表

Source: GenBank Release 134.0 (Feb 15, 2003)  
*Homo sapiens* 55194 CDS's ( 24298072 codons )

		第 2 文字												
		T			C			A			G			
第 1 文字	T	TTT	F	17.1(0.523)	TCT	S	14.7(0.184)	TAT	Y	12.1(0.438)	TGT	C	10.0(0.451)	T
		TTC	F	20.6(0.630)	TCC	S	17.6(0.220)	TAC	Y	15.5(0.562)	TGC	C	12.2(0.550)	C
		TTA	L	7.5(0.0746)	TCA	S	12.0(0.150)	TAA	終	0.7(0.25)	TGA	終	1.5(0.536)	A
		TTG	L	12.6(0.125)	TCG	S	4.4(0.0551)	TAG	終	0.6(0.214)	TGG	W	12.7(1.00)	G
	C	CTT	L	13.0(0.129)	CCT	P	17.3(0.284)	CAT	H	10.5(0.412)	CGT	R	4.6(0.0820)	T
		CTC	L	19.8(0.197)	CCC	P	20.1(0.330)	CAC	H	15.0(0.588)	CGC	R	10.7(0.191)	C
		CTA	L	7.8(0.0776)	CCA	P	16.7(0.274)	CAA	Q	12.0(0.26)	CGA	R	6.3(0.112)	A
		CTG	L	39.8(0.396)	CCG	P	6.9(0.113)	CAG	Q	34.1(0.740)	CGG	R	11.6(0.207)	G
	A	ATT	I	16.1(0.355)	ACT	T	13.0(0.243)	AAT	N	16.7(0.461)	AGT	S	11.9(0.179)	T
		ATC	I	21.6(0.476)	ACC	T	19.4(0.362)	AAC	N	19.5(0.539)	AGC	S	19.3(0.242)	C
		ATA	I	7.7(0.170)	ACA	T	15.1(0.282)	AAA	K	24.1(0.248)	AGA	R	11.5(0.205)	A
		ATG	M	22.2(1.00)	ACG	T	6.1(0.114)	AAG	K	32.2(0.572)	AGG	R	11.4(0.203)	G
	G	GTT	V	11.0(0.180)	GCT	A	18.6(0.234)	GAT	D	21.9(0.461)	GGT	G	10.8(0.164)	T
		GTC	V	14.6(0.239)	GCC	A	28.4(0.402)	GAC	D	25.6(0.539)	GGC	G	22.5(0.341)	C
		GTA	V	7.2(0.118)	GCA	A	16.1(0.228)	GAA	E	29.0(0.421)	GGA	G	16.4(0.249)	A
		GTG	V	28.4(0.464)	GCG	A	7.5(0.106)	GAG	E	39.9(0.579)	GGG	G	16.3(0.247)	G

第 3 文字	T
	C
	A
	G
	T
	C
	A
	G
	T
	C
	A
	G
	T
	C
	A
	G

数値は 1000 に対する使用頻度

( )内は、同一アミノ酸内での出現頻度(こちらは per 1 の値)

【0168】



【表 2】

各コドン中の一塩基置換により引き起こされる、コードされるアミノ酸変化

G	original code	第 1 文字の置換	使用頻度	第 2 文字の置換	使用頻度
	GGT	TGT: C	110.0	GTT: V	11.0
		CGT: R	4.6	GCT: A	18.6
		AGT: S	11.9	GAT: D	21.9
	GGC	TGC: C	12.2	GTC: V	14.6
		CGC: R	10.7	GCC: A	28.4
		AGC: S	19.3	GAC: D	25.6
	GGA	TGA: 終	1.5	GTA: V	7.2
		CGA: R	6.3	GCA: A	16.1
		AGA: R	11.5	GAA: E	29.0
	GGG	TGG: W	12.7	GTG: V	28.4
		CGG: R	11.6	GCG: A	7.5
		AGG: R	11.4	GAG: E	39.9
	C R S 終 W V A D E				
	original code	第 1 文字の置換	使用頻度	第 2 文字の置換	使用頻度
A	GCT	TCT: S	14.7	GTT: V	11.0
		CCT: P	17.3	GAT: D	21.9
		ACT: T	13.0	GGT: G	10.8
	GCC	TCC: S	17.6	GTC: V	14.6
		CCC: P	20.1	GAC: D	25.6
		ACC: T	19.4	GGC: G	22.5
	GCA	TCA: S	12.0	GTA: V	7.2
		CCA: P	16.7	GAA: E	29.0
		ACA: T	15.1	GGA: G	16.4
	GCG	TCG: S	4.4	GTG: V	28.4
		CCG: P	6.9	GAG: E	39.9
		ACG: T	6.1	GGG: G	16.3
	S P T V D G E				

【0169】

【表 3】

P	original code	第 1 文字の置換	使用頻度	第 2 文字の置換	使用頻度
	CCT	TCT: S	14.7	CTT: L	13.0
		ACT: T	13.0	CAT: H	10.5
		GCT: A	18.6	CGT: R	4.6
	CCC	TCC: S	17.6	CTC: L	19.8
		ACC: T	19.4	CAC: H	15.0
		GCC: A	28.4	CGC: R	10.7
	CCA	TCA: S	12.0	CTA: L	7.8
		ACA: T	15.1	CAA: Q	12.0
		GCA: A	16.1	CGA: R	6.3
	CCG	TCG: S	4.4	CTG: L	39.8
		ACG: T	6.1	CAG: Q	34.1
		GCG: A	7.5	CGG: R	11.6
	S T A L H R Q				
	original code	第 1 文字の置換	使用頻度	第 2 文字の置換	使用頻度
V	GTT	TTT: F	17.0	GCT: A	18.6
		CTT: L	13.0	GAT: D	21.9
		ATT: I	16.1	GGT: G	10.8
	GTC	TTC: F	20.6	GCC: A	28.4
		CTC: L	19.8	GAC: D	25.6
		ATC: I	21.6	GGC: G	22.5
	GTA	TTA: F	7.5	GCA: A	16.1
		CTA: L	7.8	GAA: E	29.0
		ATA: I	7.7	GGA: G	16.4
	GTG	TTG: F	12.6	GCG: A	7.5
		CTG: L	39.8	GAG: E	39.9
		ATG: M	22.2	GGG: G	16.3
	F L I M A O G E				

【0170】



【表 4】

T	original code	第 1 文字の置換	使用頻度	第 2 文字の置換	使用頻度
	ACT	TCT: S	14.7	ATT: I	16.1
		CCT: P	17.3	AAT: N	16.7
		GCT: A	18.6	AGT: S	11.9
	ACC	TCC: S	17.6	ATC: I	26.6
		CCC: P	20.1	AAC: N	19.5
		GCC: A	19.4	AGC: S	19.3
	ACA	TCA: S	12.0	ATA: I	7.7
		CCA: P	16.7	AAA: K	24.1
		GCA: A	16.1	AGA: R	11.5
	ACG	TCG: S	4.4	ATG: M	22.2
		CCG: P	6.9	AAG: K	32.2
		GCG: A	7.5	AGG: R	11.4
	S P A I N K R M				

【0 1 7 1】

【表 5】

C	original code	第 1 文字の置換	使用頻度	第 2 文字の置換	使用頻度
	TGT	CGT: R	4.6	TTT: F	17.1
		AGT: S	11.9	TCT: S	14.7
		GGT: G	10.8	TAT: Y	12.1
	TGC	CGC: R	10.7	TTC: F	20.6
		AGC: S	19.3	TCC: S	17.6
		GGC: G	22.5	TAC: Y	15.5
		TAG: 終	1.5	TGG: W	12.7
	R S G F Y W 終				
D	original code	第 1 文字の置換	使用頻度	第 2 文字の置換	使用頻度
	GAT	TAT: Y	12.1	GTT: V	11.0
		CAT: H	10.5	GCT: A	18.6
		AAT: N	16.7	GGT: G	10.8
	GAC	TAC: Y	15.5	GTC: V	14.6
		CAC: H	15.0	GCC: A	28.4
		AAC: N	19.5	GGC: G	22.5
		GAA: E	29.0	GAG: E	39.9
	Y H N E V A G				
N	original code	第 1 文字の置換	使用頻度	第 2 文字の置換	使用頻度
	AAT	TAT: Y	12.1	ATT: I	16.1
		CAT: H	10.5	ACT: T	13.0
		GAT: D	21.9	AGT: S	11.9
	AAC	TAC: Y	15.5	ATC: I	21.6
		CAT: H	15.0	ACC: T	19.4
		GAC: D	25.6	AGC: S	19.3
		AAA: K	24.1	AAG: K	32.2
	Y H D N K I T S				

【0172】

【表 6】

E	original code	第 1 文字の置換	使用頻度	第 2 文字の置換	使用頻度
	GAA	TAA: 終	0.7	GTA: V	7.2
		CAA: Q	12.0	GCA: A	16.1
		AAA: K	24.1	GGA: G	16.4
	GAG	TAG: 終	0.6	GTG: V	28.4
		CAG: Q	34.1	GCG: A	7.5
		AAG: K	32.2	GGG: G	16.3
		GAT: D	21.9	GAC: D	25.6
	Q K D V A G 終				
K	original code	第 1 文字の置換	使用頻度	第 2 文字の置換	使用頻度
	AAA	TAA: 終	0.7	ATA: I	7.7
		CAA: Q	12.0	ACA: A	15.1
		GAA: E	29.0	AGA: R	11.5
	AAG	TAG: 終	0.6	ATG: M	22.2
		CAG: Q	34.1	ACG: T	6.1
		GAG: E	39.9	AGG: R	11.4
		AAT: N	16.7	AAC: N	19.5
	Q E N I T R M 終				
Q	original code	第 1 文字の置換	使用頻度	第 2 文字の置換	使用頻度
	CAA	TAA: 終	0.7	CTA: L	7.8
		AAA: K	24.1	CCA: P	16.7
		GAA: E	29.0	CGA: R	6.3
	CAG	TAG: 終	0.6	CTG: L	39.8
		AAG: K	32.2	CCG: P	6.9
		GAG: E	39.9	CGG: R	11.6
		CAT: H	10.5	CAC: H	15.0
	K E H L P R 終				

【0173】



【表 7】

H	original code	第 1 文字の置換	使用頻度	第 2 文字の置換	使用頻度
	CAT	TAT: Y	12.1	CTT: L	13.0
		AAT: N	16.7	CCT: P	17.3
		GAT: D	21.9	CGT: R	4.6
	CAC	TAC: Y	15.5	CTC: L	19.8
		AAC: N	19.5	CCC: P	20.1
		GAC: D	25.6	CGC: R	10.7
		CAA: Q	12.0	CAG: Q	34.1
	Y N D Q P R L				
F	original code	第 1 文字の置換	使用頻度	第 2 文字の置換	使用頻度
	TTT	CTT: L	13.0	TAT: Y	14.7
		ATT: I	16.1	TCT: S	12.1
		GTT: V	11.0	TGT: C	10.0
	TTC	CTC: L	19.8	TAC: Y	17.6
		ATC: I	21.6	TCC: S	15.5
		GTC: V	14.6	TGC: C	12.2
		TTA: L	7.5	TTG: L	12.6
	L I V S Y C				
Y	original code	第 1 文字の置換	使用頻度	第 2 文字の置換	使用頻度
	TAT	CAT: H	10.5	TTT: F	17.1
		AAT: N	16.7	TCT: S	14.7
		GAT: D	21.9	TGT: C	10.0
	TAC	CAC: H	15.0	TTC: F	20.6
		AAC: N	19.5	TCC: S	17.6
		GAC: D	25.6	TGC: C	12.2
		TAA: 終	0.7	TAG: 終	0.6
	H N D F S C 終				

【0174】

【表 8】

S	original code	第 1 文字の置換	使用頻度	第 2 文字の置換	使用頻度
	TCT	CCT: P	17.3	TTT: F	17.1
		ACT: T	13.0	TAT: Y	12.1
		GCT: G	18.6	TGT: C	10.0
	TCC	CCC: P	20.1	TTC: F	20.6
		ACC: T	19.4	TAC: Y	15.5
		GCC: A	28.4	TGC: C	21.2
	TCA	CCA: P	16.7	TTA: L	7.5
		ACA: T	15.1	TAA: 終	0.7
		GCA: A	16.1	TGA: 終	1.5
	TCG	CCG: P	6.9	TTG: L	12.6
		ACG: T	6.1	TAG: 終	0.6
		GCG: A	7.5	TGG: W	12.7
	AGT	TGT: C	10.0	ATT: I	16.1
		CGT: R	4.6	ACT: T	13.0
		GGT: G	10.8	AAT: N	16.7
	AGC	TGC: C	12.2	ATC: I	21.6
		CGC: R	10.7	ACC: T	19.4
		GGC: G	22.5	AAC: N	19.5
		AGA: R	11.5	AGG: R	11.4
	P T G A C R F Y L W I N 終				

【0175】

【表 9】

original code	第 1 文字の置換		第 2 文字の置換	
		使用頻度		使用頻度
CTT	TTT: F	17.1	CCT: P	17.3
	ATT: I	16.1	CAT: H	10.5
	GTT: V	11.0	CGT: R	4.6
CTC	TTC: F	20.6	CCC: P	20.1
	ATC: I	21.6	CAC: H	15.0
	GTC: V	14.6	CGC: R	10.7
CTA	TTA: L	7.5	CCA: P	16.7
	ATA: I	7.7	CAA: Q	12.0
	GTA: V	7.2	AGA: R	6.3
CTG	TTG: L	12.6	CCG: P	6.9
	ATG: M	22.2	CAG: Q	34.1
	GTG: V	28.4	CGG: R	11.6
TTA	CTA: L	7.8	TCA: S	12.0
	ATA: I	7.7	TAA: 終	0.7
	GTA: V	7.2	TGA: 終	1.5
TTG	CTG: L	39.8	TCG: S	4.4
	ATG: M	22.2	TAG: 終	0.6
	GTG: V	28.4	TGG: W	12.7
	TTT: F	17.1	TTC: F	20.6
F I V (L) M P H R Q S W 終				

【 0 1 7 6 】



【表 1 0】

R	original code	第 1 文字の置換	使用頻度	第 2 文字の置換	使用頻度
	CGT	TGT: C	10.0	CTT: L	13.0
		AGT: S	11.9	CCT: P	17.3
		GGT: G	10.8	CAT: H	10.5
	CGC	TGC: C	12.2	CTC: L	19.8
		AGC: S	19.3	CCC: P	20.1
		GGC: G	22.5	CAC: H	15.0
	CGA	TGA: 終	1.5	CTA: L	7.8
		AGA: T	11.5	CCA: P	16.7
		GGA: G	16.4	CAA: Q	12.0
	CGG	TGG: W	12.7	CTG: L	39.8
		AGG: R	11.4	CCG: P	6.9
		GGG: G	16.3	CAG: Q	34.1
	AGA	TGA: 終	1.5	ATA: I	7.7
		CGA: R	6.3	ACA: T	15.1
		GGA: G	16.4	AAA: K	24.1
	AGG	TGG: W	12.7	ATG: M	22.2
		CGG: R	11.6	ACG: T	6.1
		GGG: G	16.3	AAG: K	32.2
		AGT: S	11.9	AGC: S	19.3
	C S G T W (R) L P Q I K M H 終				

【 0 1 7 7】

【表 1 1】

M	original code		使用頻度		使用頻度
	ATG	TTG: L	12.6	AGG: R	11.4
		CTG: L	39.8	ATT: I	16.1
		GTG: V	28.4	ATC: I	21.6
		ACG: T	6.1	ATA: I	7.7
		AAG: K	32.2		
	L V T K R I				

【 0 1 7 8】

【表 1 2】

W	original code		使用頻度		使用頻度
	TGG	CGG: R	411.6	TAG: 終	0.6
		AGG: R	11.4	TGT: C	10.0
		GGG: G	16.3	TGC: C	12.2
		TTG: L	12.6	TGA: 終	1.5
		TCG: S	4.4		
	R G L S C 終				

【0 1 7 9】

【表 1 3】

I	original code	第 1 文字の置換	使用頻度	第 2 文字の置換	使用頻度
	ATT	TTT: F	17.1	ACT T	13.0
		CTT: L	13.0	AAT: N	16.7
		GTT: V	11.0	AGT: S	11.9
	ATC	TTC: F	20.6	ACC: T	19.4
		CTC: L	19.8	AAC: N	19.5
		GTC: V	14.6	AGC: S	19.3
	ATA	TTA: L	7.5	ACA: T	15.1
		CTA: L	7.8	AAA: K	24.1
		GTA: V	7.2	AGA: R	11.5
		ATG: M	22.2		
	F L V M T N S K R				

【0 1 8 0】

【表 1 4】

	G	A	S	P	V	T	C	L	I	D	N	E	K	Q	M	H	F	Y	R	W	終	
G		●	●		●		●			●		●							●	●	●	G
A	●		●	●	●	●				●		●										A
S	●	●		●		●	●	●	●		●						●	●	●	●	●	S
P		●	●			●		●						●		●			●			P
V	●	●						●	●	●		●			●		●					V
T		●	●	●					●		●		●		●				●			T
C	●		●														●	●	●	●	●	C
L			●	●	●				●					●	●	●	●		●	●	●	L
I			●		●	●		●			●		●		●		●		●			I
D	●	●			●						●	●				●		●				D
N			●			●			●	●			●			●		●				N
E	●	●			●					●			●	●							●	E
K						●			●		●		●	●					●		●	K
Q				●				●				●	●			●			●			Q
M					●	●		●	●				●						●			M
H				●				●		●	●			●				●	●			H
F			●		●		●	●	●									●				F
Y			●				●			●	●					●	●				●	Y
R	●		●	●		●	●	●	●				●	●	●	●				●	●	R
W	●		●				●	●											●		●	W
終	●		●				●	●				●	●	●				●	●	●		終
	G	A	S	P	V	T	C	L	I	D	N	E	K	Q	M	H	F	Y	R	W	終	

【 0 1 8 1】



【表 15】

アミノ酸間の変換を引き起こす、塩基変異の最小ステップ数

置換後のアミノ酸残基

置換前のアミノ酸残基

	G	A	S	P	V	T	C	L	I	D	N	E	K	Q	M	H	F	Y	R	W	
G	0	1	1	2	1	2	1	2	2	1	2	1	2	2	2	2	2	2	1	1	G
A	1	0	1	1	1	1	2	2	2	1	2	1	2	2	2	2	2	2	2	2	A
S	1	1	0	1	2	1	1	1	1	2	1	2	2	2	2	2	1	1	1	1	S
P	2	1	1	0	2	1	2	1	2	2	2	2	2	1	2	1	2	2	1	2	P
V	1	1	2	2	0	2	2	1	1	1	2	1	2	2	1	2	1	2	2	2	V
T	2	1	1	1	2	0	2	2	1	2	1	2	1	2	1	2	2	2	1	2	T
C	1	2	1	2	2	2	0	2	2	2	2	3	3	3	3	2	1	1	1	1	C
L	2	2	1	1	1	2	2	0	1	2	2	2	2	1	1	1	1	2	1	1	L
I	2	2	1	2	1	1	2	1	0	2	1	2	1	2	1	2	1	2	1	3	I
D	1	1	2	2	1	2	2	2	2	0	1	1	2	2	3	1	2	1	2	3	D
N	2	2	1	2	2	1	2	2	1	1	0	2	1	2	2	1	2	1	2	3	N
E	1	1	2	2	1	2	3	2	2	1	2	0	1	1	2	2	3	2	2	2	E
K	2	2	2	2	2	1	3	2	1	2	1	1	0	1	1	2	3	2	1	2	K
Q	2	2	2	1	2	2	3	1	2	2	2	1	1	0	2	1	3	2	1	2	Q
M	2	2	2	2	1	1	3	1	1	3	2	2	1	2	0	3	2	3	1	2	M
H	2	2	2	1	2	2	2	1	2	1	1	2	2	1	3	0	2	1	1	3	H
F	2	2	1	2	1	2	1	1	1	2	2	3	3	3	2	2	0	1	2	2	F
Y	2	2	1	2	2	2	1	2	2	1	1	2	2	2	3	1	1	0	2	2	Y
R	1	2	1	1	2	1	1	1	1	2	2	2	1	1	1	1	2	2	0	1	R
W	1	2	1	2	2	2	1	1	3	3	3	2	2	2	2	3	2	2	1	0	W
	G	A	S	P	V	T	C	L	I	D	N	E	K	Q	M	H	F	Y	R	W	

【0182】

【表16】

各分子量変化を引き起こすアミノ酸置換

<i>d</i>	置換の種類	<i>d</i>	<i>d</i>	置換の種類	<i>d</i>	<i>d</i>	置換の種類	<i>d</i>
1	LN, <u>IN</u> , <u>ND</u> , ( <u>KE</u> ), <u>QE</u>	-1	27	<u>SN</u> , ( <u>TK</u> ), <u>TQ</u> , ( <u>ER</u> )	-	58	<u>GD</u> , <u>AE</u> , <u>QW</u> , ( <u>KW</u> )	-58
2	PV, VT, TC, LD, ID, EM	-2	28	<u>AV</u> , SD, TE, CM, ( <u>KR</u> ), ( <u>QR</u> )	-	59	( <u>PR</u> )	-59
3	( <u>KM</u> ), QM	-3	29	(VK), <u>VQ</u>	-	60	AM, <u>SF</u> , <u>CY</u>	-60
4	<u>PT</u> , VC	-4	30	<u>GS</u> , <u>AT</u> , <u>VE</u> , <u>TM</u> , ( <u>RW</u> )	-	62	TY	-62
6	PC, MH	-6	31	(PK), <u>PQ</u>	-	64	VY	-64
7	(RY)	-7	32	AC, PE, <u>VM</u> , DF, FY	-	66	AH, PY	-66
8	EH	-8	33	NF	-	69	( <u>SR</u> )	-69
9	<u>QH</u> , (FR), (KH)	-9	34	PM, CH, <u>LF</u> , <u>IF</u> , EY	-	71	GK, GQ, DW	-71
10	<u>SP</u> , CL, CI, HF	-	35	(KY), QY	-	72	<u>GE</u> , NW	-72
11	CN	-	36	TH, FW	-	73	<u>LW</u> , IW	-73
12	SV, TL, <u>TL</u> , CD	-	38	VH	-	74	GM	-74
13	<u>TN</u> , (DK), DQ	-	40	GP, <u>PH</u>	-	76	AF, <u>SY</u>	-76
14	GA, ST, VL, VI, TD, DE, (NK), NQ	-	41	(SK), SQ, (DR)	-	80	GH	-80
15	VN, (LK), <u>LQ</u> , ( <u>IK</u> ), IQ, NE	-	42	<u>GV</u> , AL, AI, SE, (NR)	-	83	<u>CW</u>	-83
16	<u>AS</u> , <u>SC</u> , <u>PL</u> , <u>PI</u> , <u>VD</u> , <u>LE</u> , IE, DM, MF, <u>FY</u>	-	43	AN, ( <u>LR</u> ), ( <u>IR</u> )	-	85	(AR), TW	-85
17	PN, NM	-	44	GT, <u>AD</u> , SM, <u>CF</u>	-	87	VW	-87
18	PD, LM, IM	-	46	<u>GC</u> , TF	-	89	PW	-89
19	EF	-	48	<u>VE</u> , <u>DY</u>	-	90	GF	-90
22	DH	-	50	<u>NY</u> , HW	-	92	AY	-92
23	NH, YW	-	53	SH, PF, LY, IY, ( <u>CR</u> )	-	99	<u>SW</u> , ( <u>GR</u> )	-99
24	<u>LH</u> , IH	-	55	( <u>TR</u> )	-	106	GY	-
25	(CK), CQ, ( <u>MR</u> )	-	56	GL, GI	-	115	AW	-
26	<u>AP</u> , <u>SL</u> , <u>SI</u> , CE, <u>HY</u>	-	57	GN, (AK), AQ, (VR), EW	-	129	<u>GW</u>	-

※この表で、"XY"はアミノ酸置換  $X \leftrightarrow Y$  を表す。数値は、正の数値は  $X \rightarrow Y$  の「右読み」の置換、負の数値は  $X \leftarrow Y$  の「左読み」の置換での質量差である。

※下線部は一塩基置換によって生じうるアミノ酸置換である。

※ ( ) は、トリプシンによる切断部位；アルギニン、リシンと、他のアミノ酸との間のアミノ酸置換を示す。

【0183】

【表 17】

質量差	アミノ酸置換の種類	質量差
1	ND, IN, QE	-1
16	PL, AS, SC, VD, FY	-16
26	SL, HY, AP, SI	-26
30	TM, GS, AT, VE	-30
34	LF, IF	-34
44	AD, CF	-44
48	VF, DY	-48
58	GD, AE	-58
60	SF, CY	-60

※表中、“XY”はアミノ酸置換 $X \leftrightarrow Y$ を表す。数値は、正の数値は $X \rightarrow Y$ の「右読み」の置換、負の数値は $X \leftarrow Y$ の「左読み」の置換での質量差である。

【0184】

【表 18】

質量差	transition型 塩基対置換に起因する変換	transversion型 塩基対置換に起因する変換	質量差
1	ND	IN, QE	-1
16	PL	AS>SC, VD, FY	-16
26	SL, HY	SI>AP	-26
30	TM>GS, AT	VE	-30
34	LF	IF	-34
44		AD>CF	-44
48		VF, DY	-48
58		GD, AE	-58
60	SF>CY		-60

※表中、“XY”はアミノ酸置換 $X \leftrightarrow Y$ を表す。数値は、正の数値は $X \rightarrow Y$ の「右読み」の置換、負の数値は $X \leftarrow Y$ の「左読み」の置換での質量差である。

【0185】

本発明にかかる解析方法では、基本的には、解析対象タンパク質に対して、そのペプチド鎖を部位特異的なタンパク質分解処理により断片化した、ペプチド断片を質量分析し、かかるペプチド断片の分子量を、対応する一価の「親の陽イオン種」の分子量 ( $M+H/Z$ ;  $Z=1$ )、または一価の「親の陰イオン種」の分子量 ( $M-H/Z$ ;  $Z=1$ ) として、質量分析法で測定される結果に基づき、データ・ベース上に収録される公知タンパク質との異同を判定する手法を採用している。より具体的には、想定されるアミノ酸配列のペプチド断片が示す分子量と、実測されるペプチド断片の分子量との比較を行うため、利用する質量分析法のイオン化過程で、ペプチド断片を構成するアミノ酸残基から、一部に原子団の欠落を生じない条件での測定により適する、Time-of-Flight型質量分析装置、例えば、MALDI-TOF-MS装置などを利用することが好ましい。また、二次的質量分析結果として、MS/MS質量分析法により、「親の陽イオン種」または「親の陰イオン種」について、その断片化により生成する各種の娘イオン種の分子量の測定結果を利用するが、その際にも、Time-of-Flight型質量分析装置、例えば、MALDI-TOF-MS装置などを利用して分離したイオン種について、更に、電子線照射により、かかるイオン種から生成する二次イオン種の質量を分析する、TOF-SIMS法などの、MS/MS法を利用することで、各ペプチド断片の部分構成に関する情報も利用可能となっている。例えば、これらMS/MS法を利用することで、場合によっては、ペプチド断片のN末端、およびC末端配列を特定できるものもある。

【0186】



一方、ペプチド断片化に利用する、部位特異的なタンパク質分解処理の手段としては、プロテアーゼによるペプチド断片化処理が利用可能である。好適に利用可能なプロテアーゼとしては、リシンやアルギニン残基のC末端側ペプチド結合を開裂するトリプシン、グルタミン酸残基のC末端側ペプチド結合を開裂するV8酵素、ロイシン、イソロイシン、バリン、フェニルアラニン残基のN末端側ペプチド結合を開裂するサーモリシンなどの、汎用されるペプチド断片化処理用のプロテアーゼを挙げることができる。

#### 【0187】

また、部位特異的なタンパク質分解処理は、前記切断アミノ酸配列部位の特異性を有するプロテアーゼ消化による他、例えば、メチオニン残基のC末側アミド結合における開裂に特異性を有するCNBrなどの化学的な試薬を利用する切断手法を利用することも可能である。

#### 【0188】

かかるプロテアーゼ消化、あるいは、化学的な切断手法を適用して、長いアミノ酸長のペプチド鎖より、得られるペプチド断片複数を、利用する質量分析法に応じて、所望の質量精度を達成する上で、好ましいアミノ酸長の範囲とすることが望ましい。すなわち、解析対象タンパク質から調製されるペプチド断片複数名全てについて、その「親の陽イオン種」または「親の陰イオン種」の上では、かかるプロテアーゼ消化、あるいは、化学的な切断における切断部位が、例えば、100アミノ酸当たり、15～2個程度、好ましくは、10～3個程度含有されていることが望ましい。前記の頻度で切断部位が存在していると、得られるペプチド断片の平均的なアミノ酸長は、7～50アミノ酸、好ましくは、10～35アミノ酸とでき、十分な精度で測定可能なアミノ酸長範囲とできる。

#### 【0189】

また、プロテアーゼ消化等の手段を用いて、ペプチド鎖の断片化を図るが、その際、一旦還元したCys側鎖上のスルファニル基(—SH)から再び、Cys—Cys結合が再生されることを防止する目的で、該直鎖状のペプチド鎖に対して、Cys側鎖上のスルファニル基(—SH)に対する選択的な保護基の導入を行うこともできる。例えば、ここでCys側鎖上のスルファニル基(—SH)に対して、選択的なカルボキシメチル化やピリジルエチル化などを施し、予めその保護を行う。なお、これらCys側鎖上に選択的に導入された保護基は、質量分析において、Cysの存在を確認する際、その標識原子団として利用することも可能である。

#### 【0190】

また、本発明にかかる質量分析法を利用するタンパク質の同定方法においては、解析対象タンパク質について、予め、切断部位に特異性を有するプロテアーゼ、例えば、トリプシンなどを利用して酵素消化した際、生成するペプチド断片個々の分子量を質量分析法で特定し、この一次質量分析情報に基づき、公知タンパク質において、同様のペプチド断片化を行った際に生成されると推定されるペプチド断片の推定分子量をデータ・ベース上に収録されている(推断)完全長アミノ酸配列の配列情報より算定し、実測されたペプチド断片個々の分子量と対比することで、同定の候補を選別している。一方、ペプチド・マス・フィンガープリント(PMF)法と称される手法においては、通常、参照すべきペプチド断片の分子量として、公知タンパク質に関して、予め、切断部位に特異性を有するプロテアーゼ、例えば、トリプシンなどを利用して酵素消化した際、生成するペプチド断片個々の分子量実測値が判明している場合、単離された解析対象タンパク質を、同じ酵素消化によってペプチド断片化し、質量分析法を利用してペプチド断片の各分子量を測定し、データ・ベースに収録されているペプチド断片個々の分子量と対比することで、同一性の検証を行うが、本発明では、実際に、公知タンパク質におけるペプチド断片個々の分子量実測値が入手できない場合へも、このペプチド・マス・フィンガープリント(PMF)法による同定法を拡張する際、その同定の候補の確度を高く維持する手段となっている。

#### 【0191】

具体的には、対照すべき公知タンパク質と比較した際、解析対象タンパク質は、翻訳後修飾に相違を有する、スプライシング変異体に相当する、また、「一塩基多型」に起因す

る、少数のアミノ酸置換を示す際、上述する一次対比操作に基づく、同定の候補である一次候補の公知タンパク質の選別に加えて、これら種々の要因に由来する変異の有無を判定する二次対比操作をも併用することで、その同定の候補の確度を高く維持する手段となっている。

#### 【0192】

以下に、上記の二次対比操作において、実施される未同定の該解析対象タンパク質由来実測ペプチド断片に対する個別的な解析手順の一例を、より具体的に説明する。

#### 【0193】

この実施態様においては、

該解析対象タンパク質に関する、MSデータとして、下記するペプチド断片化処理により得られるペプチド断片複数に関して、MALDI-TOF-MS法による対応する一価の「親の陽イオン種」の分子量 ( $M+H/Z$ ;  $Z=1$ )、ならびに、一価の「親の陰イオン種」の分子量 ( $M-H/Z$ ;  $Z=1$ )、更に、MALDI-TOF-MS装置により分離した「親のイオン種」について、更に、電子線照射により、かかる「親のイオン種」から生成する二次イオン種 (娘イオン種) の質量を分析する、TOF-SIMS法を利用する、MS/MS法の結果をも、利用している。

#### 【0194】

加えて、解析対象タンパク質について、上述する国際公開 WO 03/081255 A1号パンフレットに開示される「ペプチドのC末端アミノ酸配列解析方法」の手法を利用して、そのC末端アミノ酸を逐次的に切除して、ペプチドのC末端アミノ酸配列も、付加的MSデータとしている。

#### 【0195】

##### (1) ペプチド断片化処理

予め、単離処理した解析対象タンパク質に対して、還元条件: 2-スルファニルエタノール ( $HS-C_2H_4-OH$ : 2-メルカプトエタノール)、DTT (ジチオトレイトール: トレオ-1, 4-ジスルファニル-2, 3-ブタンジオール) などの還元性試薬を添加して、還元状態で電気泳動を行い、単一のスポット化を示すこと、また、その見かけの分子量 (Mapp) を確認する。

#### 【0196】

次いで、還元処理、鎖状ペプチド鎖へと変質処理を施した後、トリプシン消化により、リシン残基およびアルギニン残基のC末端側ペプチド結合を開裂することで、ペプチド断片化を行う。

#### 【0197】

##### (2) 質量分析

ペプチド断片化処理により得られるペプチド断片複数に関して、MALDI-TOF-MS法により、対応する一価の「親の陽イオン種」の分子量 ( $M+H/Z$ ;  $Z=1$ )、ならびに、一価の「親の陰イオン種」の分子量 ( $M-H/Z$ ;  $Z=1$ )、更に、MALDI-TOF-MS装置により分離した「親のイオン種」について、更に、電子線照射により、かかる「親のイオン種」から生成する二次イオン種 (娘イオン種) の質量を分析する、TOF-SIMS法を利用する、MS/MS法の結果を得る。

#### 【0198】

加えて、解析対象タンパク質について、上述する国際公開 WO 03/081255 A1号パンフレットに開示される「ペプチドのC末端アミノ酸配列解析方法」の手法を利用して、そのC末端アミノ酸を逐次的に切除して、ペプチドのC末端アミノ酸配列も、付加的MSデータとする。

#### 【0199】

従って、解析対象タンパク質由来のペプチド断片の質量実測値 (Mex) を、総数Nexのペプチド断片 {Pi:  $i=1 \sim Nex$ } について、Mex (Pi) と決定する。各ペプチド断片 {Pi:  $i=1 \sim Nex$ } についての、MS/MS法による、二次イオン種 (娘イオン種) の質量を二次MS結果とする。

## 【0 2 0 0】

(3) (推断) 完全長アミノ酸配列に基づき推定される、各公知タンパク質に対する推定ペプチド断片の推定分子量 (Mref) の算出

データ・ベース上に収録される各公知タンパク質について、その (推断) 完全長アミノ酸配列に基づき、トリプシン消化により、リシン残基およびアルギニン残基の C 末端側ペプチド結合が開裂されたと仮定した際、推定されるペプチド断片 {Prefj: j = 1 ~ Nref} について、その分子量を算定し、推定ペプチド断片の推定分子量 (Mref) のデータ・セットとする。すなわち、N 末側から、推定ペプチド断片の列: Pref1... を定義し、その推定分子量 Mref (Pref1) ... のセットを構成する。

## 【0 2 0 1】

なお、切断部位が、数アミノ酸内に近接している際には、その一方では開裂が生じないことも想定されるので、この仮定による、追加的な推定ペプチド断片の推定分子量 (Mref) のデータ・セットも作成する。

## 【0 2 0 2】

(4) 一次対比操作

各公知タンパク質について、その推定ペプチド断片の推定分子量 (Mref) のデータ・セットと、解析対象タンパク質由来のペプチド断片の質量実測値 (Mex) とを対比させ、該質量分析法の測定精度内で、一致するものを選別する。

## 【0 2 0 3】

一致すると判定される (同定される)、解析対象タンパク質由来の実測ペプチド断片数 (Nex-id)、ならびに、公知タンパク質由来の推定ペプチド断片数 (Nref-id) を決定する。同時に、一致すると判定される、解析対象タンパク質由来の実測ペプチド断片質量実測値 (Mex) の集合と、公知タンパク質由来の推定ペプチド断片、推定分子量 (Mref) の集合を特定する。未同定の解析対象タンパク質由来の実測ペプチド断片質量実測値 (Mex) の集合と、未同定の公知タンパク質由来の推定ペプチド断片、推定分子量 (Mref) の集合を特定する。

## 【0 2 0 4】

データ・ベース上の公知タンパク質について、同様の対比操作を行い、最大の公知タンパク質由来の推定ペプチド断片数 (Nref-id) を示す、公知タンパク質の群を作成し、該解析対象タンパク質に対する、同定の候補である一次候補の公知タンパク質の群とする。この段階で、一次候補の公知タンパク質の群に含まれる、公知タンパク質が一種であれば、一応、該一種類の公知タンパク質を、該解析対象タンパク質に対する、単一の同定の候補と判定する。

## 【0 2 0 5】

同時に、この一種類の公知タンパク質の (推断) 完全長アミノ酸配列上、同定された、解析対象タンパク質由来の実測ペプチド断片質量実測値 (Mex) に対応する、公知タンパク質由来の推定ペプチド断片が占めている部分を、全て特定する。

(i) その際、この公知タンパク質の (推断) 完全長アミノ酸配列上において、かかる「同定領域」が、連続するアミノ酸配列部分を構成している場合、前記「単一の同定の候補」の判定は、より高い確度と認定する。

(ii) また、かかる同定領域は、N 末側と、C 末側に二分化され、その間に、「未同定領域」が一連の領域となり、三分画化が生じている場合も、前記「単一の同定の候補」の判定は、より高い確度と認定する。

(iii) あるいは、未同定の公知タンパク質由来の推定ペプチド断片は、存在するが、未同定の解析対象タンパク質由来の実測ペプチド断片質量実測値 (Mex) は存在していない場合も、前記「単一の同定の候補」の判定は、より高い確度と認定する。

## 【0 2 0 6】

仮に、一次候補の公知タンパク質の群に含まれる、公知タンパク質が複数種存在する際には、前記の基準 (i), (ii) のいずれかを満たす候補の有無を判断し、満たすものが一種であれば、かかる一種類の公知タンパク質を、該解析対象タンパク質に対する、単



一の同定の候補と判定する。

【0 2 0 7】

この副次的判定を満たすものが無い場合には、仮に同定されている解析対象タンパク質由来の実測ペプチド断片の、MS/MS法による、二次イオン種（娘イオン種）の質量を二次MS結果と、対応する公知タンパク質由来の推定ペプチド断片が有するアミノ酸配列の間における妥当性を判断し、単一の同定の候補を判定する。さらには、必要があれば、仮に同定されている解析対象タンパク質由来実測ペプチド断片のC末端アミノ酸配列の付加的MSデータをも参照して、単一の同定の候補を判定する。

【0 2 0 8】

(5) 二次対比操作において、実施される未同定の該解析対象タンパク質由来実測ペプチド断片に対する個別的な解析

一次対比操作において、未同定とされた該解析対象タンパク質由来実測ペプチド断片に関して、下記の手順で、「単一の同定の候補」の公知タンパク質由来の未同定の推定ペプチド断片、推定分子量（Mref）と一致しなかった理由を解析する。

【0 2 0 9】

特に、

1. 翻訳後修飾
2. スプライシング
3. アミノ酸置換

の可能性に関して、個別の情報が得られる可能性がある。

【0 2 1 0】

(5-1) 翻訳後修飾

先ず、未同定とされた該解析対象タンパク質由来実測ペプチド断片について、翻訳後修飾の可能性について、検討する。

【0 2 1 1】

哺乳類において、見出される可能性がある、主な修飾、リン酸化、メチル化、アセチル化、ヒドロキシル化、ホルミル化、ピログルタミル化について、その可能性を検討する。

【0 2 1 2】

未同定の公知タンパク質由来の推定ペプチド断片、推定分子量（Mref）の集合に対して、前記の修飾が存在していると仮定して、その翻訳後修飾を有する推定ペプチド断片の推定分子量（Mref-mod）を算出して、二次的データ・セットとする。

【0 2 1 3】

各未同定の公知タンパク質由来の推定ペプチド断片について、各修飾基一つが付加された、その推定ペプチド断片の推定分子量（Mref-mod）のデータ・セットと、未同定の解析対象タンパク質由来ペプチド断片の質量実測値（Mex）とを対比させ、該質量分析法の測定精度内で、一致するものを選別する。

【0 2 1 4】

各未同定の解析対象タンパク質由来ペプチド断片の質量実測値（Mex）について、仮に、修飾基一つが付加された、推定ペプチド断片の推定分子量（Mref-mod）の一つと一致が見られた際には、その推定ペプチド断片のアミノ酸配列を参照して、その修飾基の付加が起こるアミノ酸が存在するか否かを判定する。その修飾基の付加が可能な場合には、仮に同定されている解析対象タンパク質由来の実測ペプチド断片の、MS/MS法による、二次イオン種（娘イオン種）の質量を二次MS結果と、対応する修飾基の付加がある推定ペプチド断片が有するアミノ酸配列の間における妥当性を判断する。不合理性が無い場合には、かかる未同定の解析対象タンパク質由来ペプチド断片の質量実測値（Mex）は、その修飾基一つが付加された、推定ペプチド断片に相当すると判定する。

【0 2 1 5】

同時に、かかる二次対比操作において、追加同定された、解析対象タンパク質由来ペプチド断片の質量実測値（Mex）と、公知タンパク質由来の推定ペプチド断片、推定分子量（Mref）を、上記の未同定の集合より除く。

## 【0 2 1 6】

(5-2) N末切断型タンパク質、または、C末切断型タンパク質

(4) の一次対比操作において、「単一の同定の候補」公知タンパク質の(推断)完全長アミノ酸配列上、同定された、解析対象タンパク質由来の実測ペプチド断片質量実測値(Mex)に対応する、公知タンパク質由来の推定ペプチド断片が占めている部分が、N末端側から連続しており、未同定の該解析対象タンパク質由来実測ペプチド断片が一つ残っている場合、C末切断型タンパク質の可能性が高い。あるいは、C末側から連続しており、未同定の該解析対象タンパク質由来実測ペプチド断片が一つ残っている場合、N末切断型タンパク質の可能性が高い。

## 【0 2 1 7】

C末切断型タンパク質と推定される場合には、未同定の公知タンパク質由来の推定ペプチド断片、推定分子量(Mref)の集合中、前記連続する同定領域の直後に相当する推定ペプチド断片のアミノ酸配列に基づき、逐次的に、C末のアミノ酸を除去した、一連のC末切除推定ペプチド断片の推定分子量(Mref-c-truncated)を算出して、二次的データ・セットとする。未同定の解析対象タンパク質由来ペプチド断片の質量実測値(Mex)について、前記一連のC末切除推定ペプチド断片の推定分子量(Mref-c-truncated)と対比させ、その一つと一致が見られた際には、そのC末切除推定ペプチド断片に相当すると判定する。

## 【0 2 1 8】

N末切断型タンパク質と推定される場合には、未同定の公知タンパク質由来の推定ペプチド断片、推定分子量(Mref)の集合中、前記連続する同定領域の直前に相当する推定ペプチド断片のアミノ酸配列に基づき、逐次的に、N末のアミノ酸を除去した、一連のN末切除推定ペプチド断片の推定分子量(Mref-n-truncated)を算出して、二次的データ・セットとする。未同定の解析対象タンパク質由来ペプチド断片の質量実測値(Mex)について、前記一連のN末切除推定ペプチド断片の推定分子量(Mref-n-truncated)と対比させ、その一つと一致が見られた際には、そのN末切除推定ペプチド断片に相当すると判定する。

## 【0 2 1 9】

(5-3) プロテイン・スプライシング型タンパク質またはスプライシング変異体型タンパク質

(4) の一次対比操作において、「単一の同定の候補」公知タンパク質の(推断)完全長アミノ酸配列上、同定された、解析対象タンパク質由来の実測ペプチド断片質量実測値(Mex)に対応する、公知タンパク質由来の推定ペプチド断片が占めている部分を特定した際、かかる同定領域は、N末側と、C末側に二分化され、その間に、「未同定領域」が一連の領域となり、三分画化が生じており、同時に、未同定の該解析対象タンパク質由来実測ペプチド断片が一つ残っている場合、プロテイン・スプライシング型タンパク質またはスプライシング変異体型タンパク質の可能性が高い。

## 【0 2 2 0】

その場合、「未同定領域」のN末とC末に位置する二つの、未同定の公知タンパク質由来推定ペプチド断片のアミノ酸配列を連結し、その連結部から、アミノ酸を逐次的に除去した、断片連結型の推定ペプチド断片一連の群について、その推定分子量(Mref)を算出して、二次的データ・セットとする。未同定の解析対象タンパク質由来ペプチド断片の質量実測値(Mex)について、前記一連の断片連結型推定ペプチド断片の推定分子量と対比させ、その一つと一致が見られた際には、その断片連結型推定ペプチド断片に相当すると判定する。

## 【0 2 2 1】

最終的に、仮に同定された断片連結型推定ペプチド断片のアミノ酸配列を参照し、その連結部位が、エキソンの連結部と一致すれば、スプライシング変異体型タンパク質と、一致しなければ、プロテイン・スプライシング型タンパク質と推断する。

## 【0 2 2 2】

なお、エキソンの特定に誤りがあり、その結果、参照するデータ・ベースにおける、（推断）完全長アミノ酸配列に誤謬を有する際にも、（4）の一次対比操作において、「単一の同定の候補」公知タンパク質の（推断）完全長アミノ酸配列上、同定された、解析対象タンパク質由来の実測ペプチド断片質量実測値（Mex）に対応する、公知タンパク質由来の推定ペプチド断片が占めている部分を特定した際、かかる同定領域は、N末側と、C末側に二分化され、その間に、「未同定領域」が一連の領域となり、三分画化が生じており、同時に、未同定の該解析対象タンパク質由来実測ペプチド断片が一つ残っていることがある。その際には、前記の二次対比操作において、未同定の解析対象タンパク質由来ペプチド断片の質量実測値（Mex）について、前記一連の断片連結型推定ペプチド断片の推定分子量と対比させ、その一つと一致が見られる可能性は極めて低い。逆に、一致するものが同定できない際には、かかるエキソンの特定に誤りがあることの強力な傍証と判断可能である。

#### 【0223】

（5-4） 「一塩基多型」に起因する、アミノ酸置換が生じている変異体タンパク質以上の二次対比操作を実施した後、依然として、未同定の該解析対象タンパク質由来実測ペプチド断片が存在する場合、「一塩基多型」に起因する、アミノ酸置換が生じている可能性を検討する。

#### 【0224】

具体的には、かかるペプチド断片中に、「一塩基多型」に起因する、アミノ酸置換が一つ含まれる可能性を検討する。依然として、未同定の「単一の同定の候補」公知タンパク質由来の推定ペプチド断片のうち、依然として、未同定の推定ペプチド断片、推定分子量（Mref）の集合に含まれるものについて、そのアミノ酸配列中に、一つのアミノ酸置換が生じた際に想定される、一群の推定ペプチド断片と、その推定分子量（Mref）を算定する。

#### 【0225】

まず、「一塩基多型」に起因する、一つのアミノ酸置換が生じた際に、変化する質量差について考察する。表16に示す結果に基づき、

- ・アミノ酸置換によって発生し得る質量差の集合：D
- ・一塩基置換に起因するアミノ酸置換で発生する質量差の集合：D<sub>1</sub>
- ・二塩基以上の置換に起因するアミノ酸置換で発生する質量差の集合：D<sub>2</sub>

$$D = D_1 \cup D_2$$

$$D_1 = \{\pm 1, \pm 3, \pm 4, \pm 9, \pm 10, \pm 12, \pm 13, \pm 14, \pm 16, \pm 18, \pm 19, \pm 22, \pm 23, \pm 24, \pm 25, \pm 26, \pm 27, \pm 28, \pm 30, \pm 31, \pm 32, \pm 34, \pm 40, \pm 42, \pm 43, \pm 44, \pm 46, \pm 48, \pm 49, \pm 53, \pm 55, \pm 58, \pm 59, \pm 60, \pm 69, \pm 72, \pm 73, \pm 76, \pm 83, \pm 99, \pm 129\}$$

$$D_2 = \{\pm 2, \pm 6, \pm 7, \pm 8, \pm 11, \pm 17, \pm 29, \pm 33, \pm 35, \pm 36, \pm 38, \pm 41, \pm 50, \pm 56, \pm 57, \pm 62, \pm 64, \pm 66, \pm 71, \pm 74, \pm 80, \pm 87, \pm 89, \pm 90, \pm 92, \pm 106, \pm 115\}$$

という集合を定義する。

#### 【0226】

（i） 未同定の公知タンパク質由来推定ペプチド断片のアミノ酸配列内に「一塩基多型」に起因する、一つのアミノ酸置換が生じていると仮定する。

#### 【0227】

図6に例示するように、これまでの二次対比操作における各ステップを経た後も、依然として、未同定とされている、公知タンパク質由来推定ペプチド断片の集合Pref-nf $\equiv$  {Pnf}、未同定とされている、解析対象タンパク質由来の実測ペプチド断片の集合Pex-ni $\equiv$  {Pni}を考慮する。

#### 【0228】



手順 i - 1:

未同定とされている、公知タンパク質由来推定ペプチド断片の集合  $\text{Pref-nf} \equiv \{ \text{Pnf} \}$  に属する、推定ペプチド断片  $\text{Pnf}$  の推定分子量  $\text{Mref}(\text{Pnf})$  に基づき、

各  $\text{Pnf} \in \text{Pref-nf} \equiv \{ \text{Pnf} \}$  に対して、

推定ペプチド断片中に一つのアミノ酸置換が生じたと仮定した際に、取りえる推定分子量  $\text{Mref}$  の集合を、

$\text{Cref-rep}(\text{Pnf}) = \{ (\text{Mref}(\text{Pnf}) + d) ; d \in D \}$  と定義する。

## 【0 2 2 9】

手順 i - 2:

一方、未同定とされている、解析対象タンパク質由来の実測ペプチド断片の集合  $\text{Pex-ni} \equiv \{ \text{Pni} \}$  における質量実測値 ( $\text{Mex}$ ) の集合を、

$\text{Cex-ni} = \{ \text{Mex}(\text{Pni}) ; \text{Pni} \in \text{Pex-ni} \}$  と定義する。

## 【0 2 3 0】

手順 i - 3:

各  $\text{Pnf} \in \text{Pref-nf} \equiv \{ \text{Pnf} \}$  に対して、

集合  $\text{Cref-rep}(\text{Pnf})$  と、集合  $\text{Cex-ni}$  との積集合を特定する。なお、その際、利用している質量分析法の測定精度を考慮して、実質的に一致するか否かを決定する。

## 【0 2 3 1】

(a) 積集合  $\text{Cref-rep}(\text{Pnf}) \cap \text{Cex-ni} = \phi$  (空集合である) の場合

該未同定の公知タンパク質由来推定ペプチド断片  $\text{Pnf}$  から一つのアミノ酸置換で生じたペプチド断片は、未同定とされている、解析対象タンパク質由来の実測ペプチド断片の集合中に存在していない。

## 【0 2 3 2】

(b) 積集合  $\text{Cref-rep}(\text{Pnf}) \cap \text{Cex-ni} \neq \phi$  (空集合ではない) の場合

該未同定の公知タンパク質由来推定ペプチド断片  $\text{Pnf}$  から一つのアミノ酸置換で生じたペプチド断片が、未同定とされている、解析対象タンパク質由来の実測ペプチド断片の集合中に存在している可能性がある。

## 【0 2 3 3】

その際、この積集合  $\text{Cref-rep}(\text{Pnf}) \cap \text{Cex-ni}$  を与える、アミノ酸置換によって発生し得る質量差  $d$  に関して、表 1 6 に示す結果を参照して、置換前のアミノ酸  $X$  と置換後のアミノ酸  $Y$  に関する組み合わせの群を特定する。

## 【0 2 3 4】

この群に含まれる置換前のアミノ酸  $X$  が、該未同定の公知タンパク質由来推定ペプチド断片  $\text{Pnf}$  のアミノ酸配列中に存在するか否かを検証する。

## 【0 2 3 5】

存在していない場合には、該未同定の公知タンパク質由来推定ペプチド断片  $\text{Pnf}$  から一つのアミノ酸置換で生じたペプチド断片は、未同定とされている、解析対象タンパク質由来の実測ペプチド断片の集合中に存在していない。

## 【0 2 3 6】

存在している場合には、該未同定の公知タンパク質由来推定ペプチド断片  $\text{Pnf}$  から一つのアミノ酸置換で生じたペプチド断片は、未同定とされている、解析対象タンパク質由来の実測ペプチド断片の集合中に存在している可能性が高い。

## 【0 2 3 7】

積集合  $\text{Cref-rep}(\text{Pnf}) \cap \text{Cex-ni}$  が、複数の要素を含む場合には、前記の検証を行うことで、一般に、可能性はより高いものが一つ選別される。この検証後、なお、二以上の要素が残る場合には、アミノ酸置換によって発生し得る質量差  $d$  が、集合  $D_1$  に属するか否かを検証し、集合  $D_1$  に属するものを、より一層可能性が高いものとして、選別する。

## 【0 2 3 8】

以上の手順 i - 3 における対比操作の結果、未同定とされている、解析対象タンパク質由来の実測ペプチド断片の集合  $\text{Pex-ni} \equiv \{ \text{Pni} \}$  の一つの実測ペプチド断片  $\text{Pni}$  が、複数

の、未同定の公知タンパク質由来推定ペプチド断片Pnfから一つのアミノ酸置換で生じたペプチド断片とより高い可能性がある、判定されることも、極稀に想定される。

#### 【0239】

従って、仮に解析対象タンパク質由来の実測ペプチド断片の集合 $Pex-ni \equiv \{Pni\}$ の一つの実測ペプチド断片Pniが、未同定の公知タンパク質由来推定ペプチド断片Pnfから一つのアミノ酸置換で生じたペプチド断片と判定された際には、その推定される置換を含むアミノ酸配列と、MS/MS法で得られる該実測ペプチド断片の二次的質量分析結果との対比を行い、その対応を検証する。あるいは、該実測ペプチド断片に対する、C末アミノ酸配列の解析結果との対比を行い、その対応を検証する。

#### 【0240】

なお、以上に示す  $i-1 \sim i-3$  の手順は、未同定の公知タンパク質由来推定ペプチド断片の集合 $Pref-nf \equiv \{Pnf\}$ の要素数が、未同定の解析対象タンパク質由来の実測ペプチド断片の集合 $Pex-ni \equiv \{Pni\}$ の要素数より少ない際に、適しており、逆に、未同定の解析対象タンパク質由来の実測ペプチド断片の集合 $Pex-ni \equiv \{Pni\}$ の要素数が、未同定の公知タンパク質由来推定ペプチド断片の集合 $Pref-nf \equiv \{Pnf\}$ の要素数より少ない際には、各解析対象タンパク質由来の実測ペプチド断片Pniに対して、アミノ酸置換によってかかる質量実測値(Mex)を与える可能性を有する、未同定の公知タンパク質由来推定ペプチド断片の有無を判定する手順を採用することができる。

#### 【0241】

具体的には、各 $Pni \in Pex-ni \equiv \{Pni\}$ に対して、一つのアミノ酸置換により、その質量実測値(Mex)が与えられていると仮定した際に、その置換前に推定される分子量の集合を、 $Cex-rep(Pni) = \{(Mex(Pni) - d) ; d \in D\}$ と定義する。一方、未同定とされている、公知タンパク質由来推定ペプチド断片の集合 $Pref-nf \equiv \{Pnf\}$ における推定分子量(Mref)の集合を、 $Cref-nf = \{Mref(Pnf) ; Pnf \in Pref-nf\}$ と定義する。

#### 【0242】

その後は、前記の手順  $i-3$  と、同様の対比操作を実施する。

#### 【0243】

(ii) 未同定の公知タンパク質由来推定ペプチド断片のアミノ酸配列内に「一塩基多型」に起因する、一つのアミノ酸置換が生じ、トリプシンによる切断部位が新たに生成していると仮定する。

#### 【0244】

この場合、図4に例示するように、未同定の公知タンパク質由来推定ペプチド断片から、二つの部分断片が生成していると推定する。そのうち、N末側の部分断片に着目すると、アミノ酸置換によって、置換前のアミノ酸XからリシンKまたはアルギニンRへと変換された部分断片となっている。従って、この種のN末側の部分断片が有する可能性がある分子量を推定する。同時に、対応するC末側の部分断片に関しても、その分子量を推定する。

#### 【0245】

##### 手順ii-1:

未同定とされている、公知タンパク質由来推定ペプチド断片の集合 $Pref-nf \equiv \{Pnf\}$ に属する、推定ペプチド断片Pnfのアミノ酸配列 $X_1(Pnf), \dots, X_n(Pnf)$ と、そのアミノ酸残基の式量 $m_1, \dots, m_n$ に基づき、

各 $Pnf \in Pref-nf \equiv \{Pnf\}$ に対して、

$X_k(Pnf)$ が、Kに置換された際に想定される、N末側の部分断片の推定分子量 $Mref-N(Pnf; X_k \rightarrow K) = (m_1 + \dots + m_{k-1}) + m_K + 18$ 、の一群；

$X_k(Pnf)$ が、Rに置換された際に想定される、N末側の部分断片の推定分子量 $Mref-N(Pnf; X_k \rightarrow R) = (m_1 + \dots + m_{k-1}) + m_R + 18$ 、の一群；

対応するC末側の部分断片の推定分子量 $Mref-C(Pnf; X_k \rightarrow K \text{ or } R) = (m_{k+1} + \dots + m_n) + 18$ 、の一群；

を算出する。

【0246】

これら新たに算出した、推定分子量の一群  $\{M_{\text{ref}}-N(P_n f; X_k \rightarrow K); k=1, \dots, n-1\}$ 、 $\{M_{\text{ref}}-N(P_n f; X_k \rightarrow R); k=1, \dots, n-1\}$ 、ならびに、 $\{M_{\text{ref}}-C(P_n f; X_k \rightarrow K \text{ or } R); k=1, \dots, n-1\}$  の各集合を定義する。

【0247】

手順ii-2:

一方、未同定とされている、解析対象タンパク質由来の実測ペプチド断片の集合  $P_{\text{ex-ni}} \equiv \{P_{\text{ni}}\}$  における質量実測値 ( $M_{\text{ex}}$ ) の集合を、 $C_{\text{ex-ni}} = \{M_{\text{ex}}(P_{\text{ni}}); P_{\text{ni}} \in P_{\text{ex-ni}}\}$  と定義する。

【0248】

手順ii-3:

各  $P_{\text{nf}} \in P_{\text{ref-nf}} \equiv \{P_{\text{nf}}\}$  に対して、各集合  $\{M_{\text{ref}}-N(P_n f; X_k \rightarrow K); k=1, \dots, n-1\}$ 、 $\{M_{\text{ref}}-N(P_n f; X_k \rightarrow R); k=1, \dots, n-1\}$ 、ならびに、 $\{M_{\text{ref}}-C(P_n f; X_k \rightarrow K \text{ or } R); k=1, \dots, n-1\}$  と、集合  $C_{\text{ex-ni}}$  との積集合を特定する。なお、その際、利用している質量分析法の測定精度を考慮して、実質的に一致するか否かを決定する。

【0249】

(c) 積集合  $[\{M_{\text{ref}}-N(P_n f; X_k \rightarrow K); k=1, \dots, n-1\} \cup \{M_{\text{ref}}-N(P_n f; X_k \rightarrow R); k=1, \dots, n-1\}] \cap C_{\text{ex-ni}} = \phi$  (空集合である) の場合

該未同定の公知タンパク質由来推定ペプチド断片  $P_{\text{nf}}$  から一つのアミノ酸置換で生じた、トリプシン切断部位に起因して、派生する N 末側のペプチド断片は、未同定とされている、解析対象タンパク質由来の実測ペプチド断片の集合中に存在していない。

【0250】

(d) 積集合  $[\{M_{\text{ref}}-N(P_n f; X_k \rightarrow K); k=1, \dots, n-1\} \cup \{M_{\text{ref}}-N(P_n f; X_k \rightarrow R); k=1, \dots, n-1\}] \cap C_{\text{ex-ni}} \neq \phi$  (空集合ではない) の場合

該未同定の公知タンパク質由来推定ペプチド断片  $P_{\text{nf}}$  から一つのアミノ酸置換で生じた、トリプシン切断部位に起因して、派生する N 末側のペプチド断片は、未同定とされている、解析対象タンパク質由来の実測ペプチド断片の集合中に存在している可能性がある。

【0251】

但し、かかる派生する N 末側のペプチド断片の推定分子量が、質量分析法における適正な測定領域より小さな場合など、質量実測値 ( $M_{\text{ex}}$ ) が得られていない場合も、排除できない。従って、派生する可能性のある C 末側のペプチド断片においても、同様の対比を行う。

【0252】

(e) 積集合  $\{M_{\text{ref}}-C(P_n f; X_k \rightarrow K \text{ or } R); k=1, \dots, n-1\} \cap C_{\text{ex-ni}} = \phi$  (空集合である) の場合

該未同定の公知タンパク質由来推定ペプチド断片  $P_{\text{nf}}$  から一つのアミノ酸置換で生じた、トリプシン切断部位に起因して、派生する C 末側のペプチド断片は、未同定とされている、解析対象タンパク質由来の実測ペプチド断片の集合中に存在していない。

【0253】

(f) 積集合  $\{M_{\text{ref}}-C(P_n f; X_k \rightarrow K \text{ or } R); k=1, \dots, n-1\} \cap C_{\text{ex-ni}} \neq \phi$  (空集合ではない) の場合

該未同定の公知タンパク質由来推定ペプチド断片  $P_{\text{nf}}$  から一つのアミノ酸置換で生じた、トリプシン切断部位に起因して、派生する C 末側のペプチド断片は、未同定とされている、解析対象タンパク質由来の実測ペプチド断片の集合中に存在している可能性がある。

【0254】

以上の対比操作を、未同定とされている、公知タンパク質由来推定ペプチド断片の集合



Pref-nf  $\equiv$  {Pnf} に属する、推定ペプチド断片Pnf全てについて、実施する。その際、未同定の解析対象タンパク質由来の実測ペプチド断片に関して、偶々、二つ以上の未同定の公知タンパク質由来推定ペプチド断片Pnfから派生した部分断片の可能性があると判定されることもある。その場合、その推定される部分アミノ酸配列と、MS/MS法で得られる該実測ペプチド断片の二次的質量分析結果との対比を行い、その対応を検証する。あるいは、該実測ペプチド断片に対する、C末アミノ酸配列の解析結果との対比を行い、その対応を検証する。

#### 【0255】

理想的には、未同定の公知タンパク質由来推定ペプチド断片のアミノ酸配列内に「一塩基多型」に起因する、一つのアミノ酸置換が生じ、トリプシンによる切断部位が新たに生成し、その結果、二つの派生した部分断片が存在する可能性が、(d) と (f) とで示唆される。また、場合によっては、(d) と (f) の一方でしか、その可能性が示唆されない場合がある。いずれの場合も、その推定される部分アミノ酸配列と、MS/MS法で得られる該実測ペプチド断片の二次的質量分析結果との対比を行い、その対応を検証する。あるいは、該実測ペプチド断片に対する、C末アミノ酸配列の解析結果との対比を行い、その対応を検証する。

#### 【0256】

(iii) 未同定の公知タンパク質由来推定ペプチド断片のアミノ酸配列内に「一塩基多型」に起因する、一つのアミノ酸置換が生じ、トリプシンによる切断部位が一つ消失していると仮定する。

#### 【0257】

この場合には、未同定とされている、公知タンパク質由来推定ペプチド断片の集合Pref-nf  $\equiv$  {Pnf} に属する、推定ペプチド断片Pnfのうち、二つが、公知タンパク質の（推断）完全長アミノ酸配列上、連続した位置を占めているはずである。

#### 【0258】

この互いに連続する二つの推定ペプチド断片Pnf間のトリプシン切断部位である、リシンまたはアルギニンが、他のアミノ酸へと置換された結果、切断がなされていないと想定する。

#### 【0259】

表16を参照して、リシンから、アルギニン以外の他のアミノ酸への置換した際の、質量数の変化の集合 $D_K \rightarrow$ と、アルギニンから、リシン以外の他のアミノ酸への置換した際の、質量数の変化の集合 $D_R \rightarrow$ とを、定義する。

#### 【0260】

$D_K \rightarrow = \{-71, -57, -31, -29, -27, -25, -15, -14, -13, +1, +3, +9, +19, +35, +58\}$

$D_R \rightarrow = \{-99, -85, -69, -57, -55, -53, -43, -42, -41, -27, -25, -19, -9, +7, +30\}$

#### 手順iii-1:

未同定とされている、公知タンパク質由来推定ペプチド断片の集合Pref-nf  $\equiv$  {Pnf} に属する、隣接する二つの推定ペプチド断片Pnf1とPnf2のアミノ酸配列に基づき、そのトリプシンによる切断部位のアミノ酸が、リシンか、アルギニンかが特定できる。

#### 【0261】

その際、リシンまたはアルギニンが、他のアミノ酸へと置換された結果、切断がなされていないと想定する際、その連結されたペプチド断片において、推定される分子量の群を算定する。

#### 【0262】

$\{ (M_{ref}(Pnf1) + M_{ref}(Pnf2) - 18 + d) ; d \in D_K \rightarrow \}$

$\{ (M_{ref}(Pnf1) + M_{ref}(Pnf2) - 18 + d) ; d \in D_R \rightarrow \}$

#### 手順iii-2:

一方、未同定とされている、解析対象タンパク質由来の実測ペプチド断片の集合Pex-ni

$\equiv \{P_{ni}\}$  における質量実測値 (Mex) の集合を、  
 $C_{ex-ni} = \{Mex(P_{ni}) ; P_{ni} \in P_{ex-ni}\}$  と定義する。

### 【0 2 6 3】

#### 手順iii-3:

各Pnf1とPnf2の連続した推定ペプチド断片組に対して、  
 予め、いずれかが定義されている、集合  $\{ (M_{ref}(P_{nf1}) + M_{ref}(P_{nf2}) - 18 + d) ; d \in D_K \rightarrow \}$ 、あるいは集合  $\{ (M_{ref}(P_{nf1}) + M_{ref}(P_{nf2}) - 18 + d) ; d \in D_R \rightarrow \}$  と、集合  $C_{ex-ni}$  との積集合を特定する。なお、その際、利用している質量分析法の測定精度を考慮して、実質的に一致するか否かを決定する。

### 【0 2 6 4】

(g) 積集合  $\{ (M_{ref}(P_{nf1}) + M_{ref}(P_{nf2}) - 18 + d) ; d \in D_K \rightarrow \} \cap C_{ex-ni} = \phi$  (空集合である)、あるいは、積集合  $\{ (M_{ref}(P_{nf1}) + M_{ref}(P_{nf2}) - 18 + d) ; d \in D_R \rightarrow \} \cap C_{ex-ni} = \phi$  (空集合である) の場合

該未同定の公知タンパク質由来推定ペプチド断片Pnfから一つのアミノ酸置換で生じた、トリプシン切断部位の消失に起因して、連結されているペプチド断片は、未同定とされている、解析対象タンパク質由来の実測ペプチド断片の集合中に存在していない。

### 【0 2 6 5】

(h) 積集合  $\{ (M_{ref}(P_{nf1}) + M_{ref}(P_{nf2}) - 18 + d) ; d \in D_K \rightarrow \} \cap C_{ex-ni} \neq \phi$  (空集合ではない)、あるいは、積集合  $\{ (M_{ref}(P_{nf1}) + M_{ref}(P_{nf2}) - 18 + d) ; d \in D_R \rightarrow \} \cap C_{ex-ni} \neq \phi$  (空集合ではない) の場合

該未同定の公知タンパク質由来推定ペプチド断片Pnfから一つのアミノ酸置換で生じた、トリプシン切断部位の消失に起因して、連結されているペプチド断片が、未同定とされている、解析対象タンパク質由来の実測ペプチド断片の集合中に存在している可能性がある。

### 【0 2 6 6】

その場合、その推定される部分アミノ酸配列と、MS/MS法で得られる該実測ペプチド断片の二次的質量分析結果との対比を行い、その対応を検証する。あるいは、該実測ペプチド断片に対する、C末アミノ酸配列の解析結果との対比を行い、その対応を検証する。

。

### 【0 2 6 7】

同時に、この連結するペプチド断片を与える、dの値から、表16を参照して、リシンまたはアルギニンが、どのような、他のアミノ酸へと置換されているかを特定可能である。

。

### 【0 2 6 8】

#### (5-5) de novo sequencingの活用

上述する一連の二次対比操作において、一次対比操作の際に選定された、該解析対象タンパク質に対する、単一の同定の候補である、一種類の公知タンパク質の(推断)完全長アミノ酸配列に基づき、解析対象タンパク質に由来する未同定のペプチド断片に対する、可能性の高い同定候補の推定がなされる。

### 【0 2 6 9】

その際、上述するように、推定ペプチド断片を利用するPMF法及びMS/MS分析の結果に基づき、確度の高い、有意な同定が可能ではあるものの、これら未同定のペプチド断片については、MS/MS分析で得られるフラグメント・イオン種の結果を利用して、可能な限りde novo sequenceにより得られる該未同定のペプチド断片に含まれる、部分アミノ酸配列の推定結果、さらには、該実測ペプチド断片に対する、C末アミノ酸配列の解析結果などを、各同定された配列とを重ね合わせてみることで、更に確度の高い、局所的なアミノ酸置換、修飾基の付加の可能性を検討できる。実際に、de novo sequenceの結果と、単一の同定の候補である、公知タンパク質から推定される配列との間に、部分的な相違があり、かかる相違部分が、二次対比操作により特定されたアミノ酸置換に相当することが確認されると、その同定の信頼度は一層高いものとなる。

## 【0 2 7 0】

更には、翻訳後修飾と、アミノ酸置換とが同時に生じている場合には、上記の一連の二次対比操作では、その特定はなされないが、de novo sequenceにより得られる部分アミノ酸配列の推定結果、ならびに、該実測ペプチド断片に対する、C末アミノ酸配列の解析結果をも利用すると、特定が可能となる場合もある。

## 【0 2 7 1】

逆に、例えば、誤って、質量分析に際して、「ノイズピーク」を、解析対象タンパク質由来の実測ペプチド断片のピークと判断すること、MS/MS分析に基づき、de novo sequencingを実施することで、排除できることになる。具体的には、解析対象タンパク質は、予め単離処理を施しているものの、例えば、二次元電気泳動法で分離を図った場合でも、極めて隣接しているスポットを与える他のタンパク質が、若干混入することも少なくない。これら混入する他のタンパク質の総量は少ないが、質量分析の際、イオン化効率の高いペプチド断片を生成すると、その混入タンパク質由来のペプチド断片に起因するピークを、解析対象タンパク質由来の実測ペプチド断片のピークうち、イオン化効率が低いものと誤認する可能性がある。MS/MS分析に基づき、de novo sequencingを実施することで、この種の誤認を回避することができる。

## 【0 2 7 2】

また、MALDI-TOF-MS法では、ペプチド断片由来の対応する一価の「親の陽イオン種」( $M+H/Z$ ;  $Z=1$ )、または一価の「親の陰イオン種」( $M-H/Z$ ;  $Z=1$ )が主に生成されるものの、高次イオン化されたイオン種( $Z \geq 2$ )も若干生成される。あるいは、一旦生成した一価の「親の陽イオン種」( $M+H/Z$ ;  $Z=1$ )、または一価の「親の陰イオン種」( $M-H/Z$ ;  $Z=1$ )が、フラグメンテーションを起こす「PSD (post source decay)」という現象もあり、場合によっては、このPSD現象で生じた派生イオン種のピークも観測されることもある。これら解析対象タンパク質由来のペプチド断片に起因する派生的なイオン種ピークは、通常、ピーク強度は小さいものの、ペプチド断片由来の対応する一価の「親の陽イオン種」( $M+H/Z$ ;  $Z=1$ )、または一価の「親の陰イオン種」( $M-H/Z$ ;  $Z=1$ )と混同する可能性もあるが、MS/MS分析に基づき、de novo sequencingを実施することで、この種の混同も排除することができる。

## 【0 2 7 3】

(6) 疾患に関連した翻訳後修飾、スプライシング変異体、「一塩基多型」アミノ酸置換の示唆

上述する一連の二次対比操作によって、翻訳後修飾、スプライシング変異体、「一塩基多型」アミノ酸置換の存在を示唆する判定が得られると、これらの変異と疾患との間の関連性を考察する上で、大きな指針を与えられ思考えられる。

## 【0 2 7 4】

正常者のサンプルと疾患罹患者のサンプルとについて、differential解析を行った際、同定の候補とされる公知タンパク質は同じであるものの、疾患罹患者のサンプルに由来する対象タンパク質において、翻訳後修飾、スプライシング変異体、「一塩基多型」アミノ酸置換の存在が示唆される場合、疾患特異的な翻訳後修飾、スプライシング変異体、「一塩基多型」アミノ酸置換の可能性が示唆され思考えられる。

## 【0 2 7 5】

翻訳後修飾やスプライシング変異体に関しては、多くの場合、二次元電気泳動において、互いに、二次元的に分離されたスポットとして現れるため、何らかの相違を有することは判断できるが、具体的に、その相違はどのようなものかに関して、上記の本発明にかかる同定方法における、二次対比操作によって入手される情報は、大きな価値を有する思考えられる。

## 【0 2 7 6】

因みに、仮に、スプライシング機構に異常を来たした場合、機能を失ったタンパクが発現し、そのタンパクが種々の疾患（特に難治性神経疾患）の発症に関する可能性が指摘され



ている。スプライシング障害により発症する疾患として、前頭側頭型痴呆 (tau遺伝子)、脊髄性筋萎縮症 (SMN1遺伝子)、孤発性筋萎縮性側索硬化症 (グルタミン酸トランスポーターEAAT2遺伝子) を代表として数多くの難治性神経疾患が報告されている。この種のスプライシング異常に由来するタンパク質に関しては、differential解析に拠らなくとも、本発明の方法で利用されるデータ・ベースにおいて、ゲノム遺伝子の塩基配列情報上に、正常なタンパク質のexon-intronの構造が収録されてさえいれば、前述するように、本発明の方法を利用して、その異常を示唆することができると考えられる。

【産業上の利用可能性】

【0 2 7 7】

本発明にかかる質量分析法を利用するタンパク質の同定方法は、特に、解析対象タンパク質を構成するペプチド鎖が、データ・ベースに収録されている、対応するゲノム遺伝子によってコードされる完全長アミノ酸配列を有するペプチド鎖と対比して、上述するような種々の要因に起因して、特異的な変異、修飾を有する場合においても、解析対象タンパク質の質量分析法で得られる情報に基づき、公知タンパク質に関するデータ・ベースに収録される、公知タンパク質個々に関して、該公知タンパク質を構成するペプチド鎖の完全長アミノ酸配列をコードしているゲノム遺伝子、および該完全長アミノ酸配列の翻訳を可能とする、mRNA中の読み枠の塩基配列、ならびに、その塩基配列でコードされる(推断)完全長アミノ酸配列の配列情報を参照して、該解析対象タンパク質に相当すると推定される、前記データ・ベースに収録される、公知タンパク質の一つを高い確度で選定する方法となる。従って、発現タンパク質の変異、修飾異常等が、その疾患の発症、進行と相関を有する場合などにおいて、かかる変異体タンパク質、修飾異常の詳細な解析に必要となる、対応する正常なタンパク質の特定、あるいは、対応する遺伝子の特定を高い確度で行うこと、また、変異あるいは修飾異常の有無をも高い確度で予測することを可能とする。

【図面の簡単な説明】

【0 2 7 8】

【図1】 alternative splicing過程による、同一のゲノム遺伝子上から翻訳される、二種のスプライシング変異体、ならびに、エキソン領域の同定の誤謬がある際、実際に翻訳されるペプチド鎖のコード領域を模式的に示す図である。

【図2】 プロテイン・スプライシング過程に起因する、翻訳後のペプチド鎖の部分的除去と、そのペプチド鎖の部分的除去に起因する、プロテアーゼ消化によるペプチド断片化の相違を模式的に示す図である。

【図3】 翻訳後、ペプチド鎖のC末部分の切除を受けた、C末切除型タンパク質と、完全長アミノ酸配列を有する前駆体における、プロテアーゼ消化によるペプチド断片化の相違を模式的に示す図である。

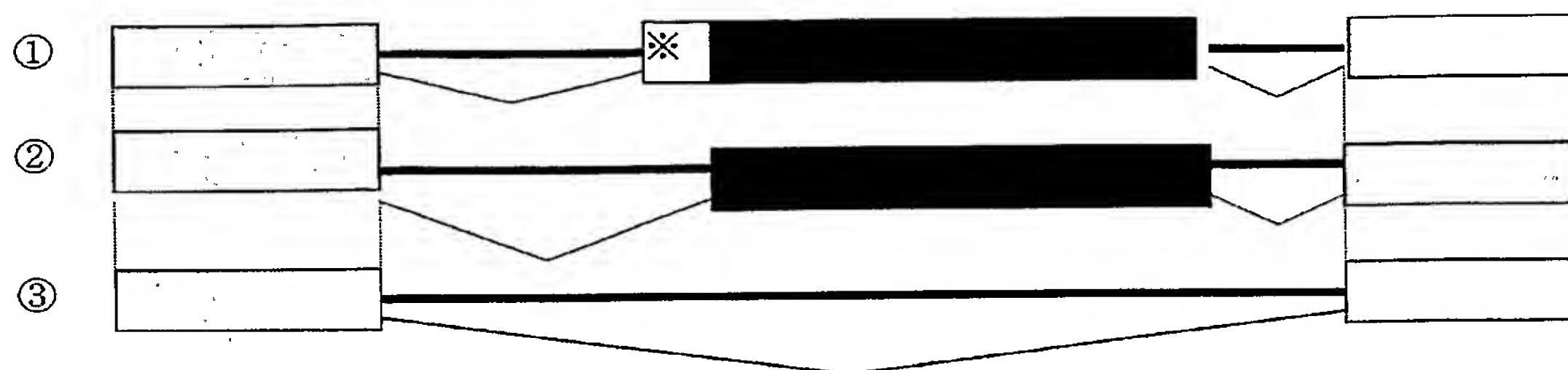
【図4】 「一塩基多型」に起因して、ペプチド断片中に切断部位が導入され、プロテアーゼ消化によって、二つに切断されたペプチド断片を派生する形態を模式的に示す図である。

【図5】 「一塩基多型」に起因して、隣接するペプチド断片間の切断部位が消失し、プロテアーゼ消化において、二つのペプチド断片部が連結されているペプチド断片が残される形態を模式的に示す図である。

【図6】 同定された解析対象タンパク質由来の実測ペプチド断片数 (Nex-id) ならびに公知タンパク質由来の推定ペプチド断片数 (Nref-id) と、未同定の解析対象タンパク質由来の実測ペプチド断片数 (Nex-ni) と公知タンパク質由来の推定ペプチド断片数 (Nref-nf) を模式的に示す図である。

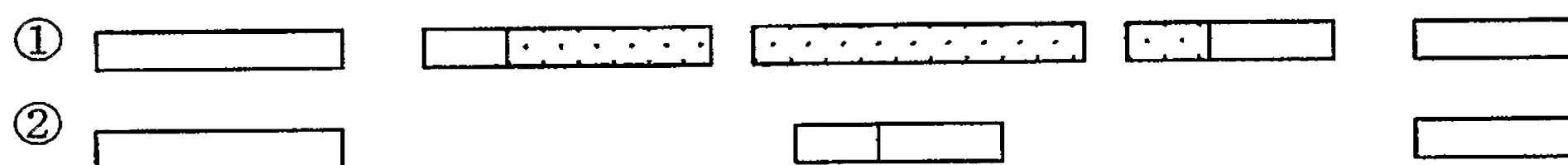
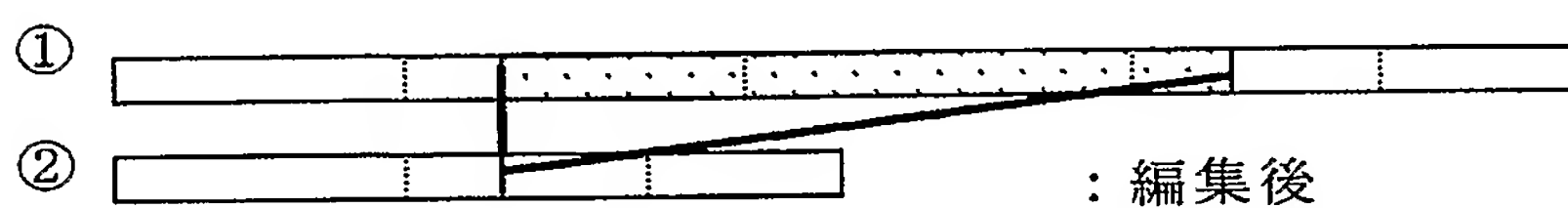
【書類名】 図面

【図 1】



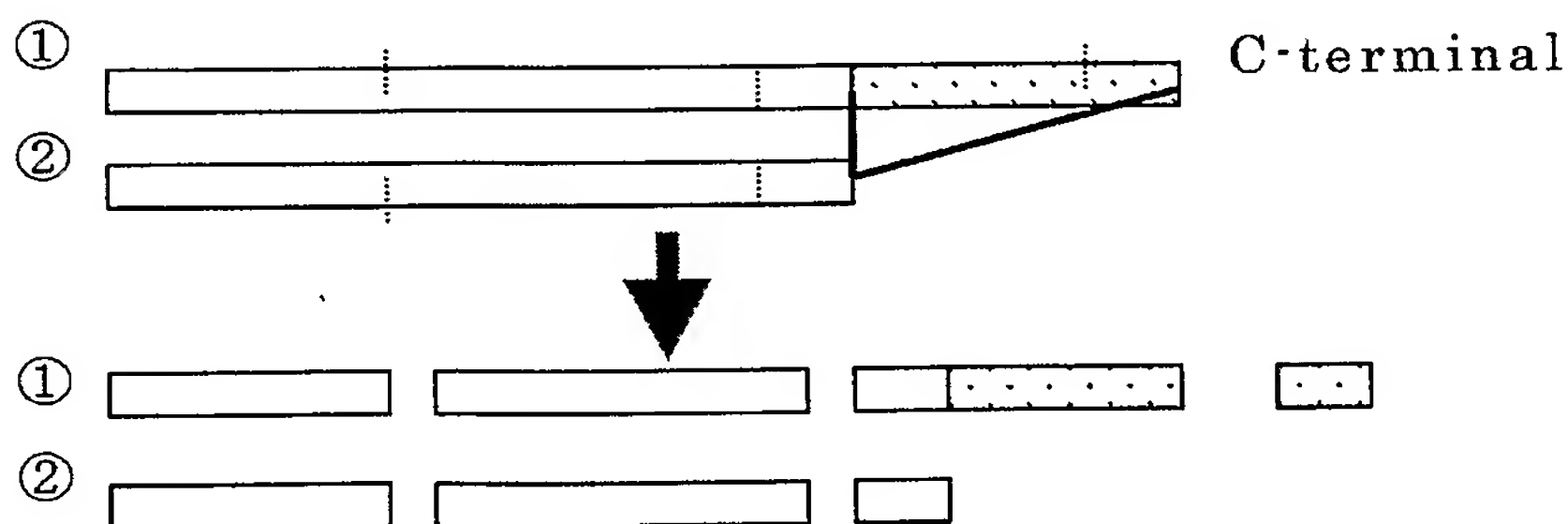
※部分の塩基対数は3の倍数ではないものとする

【図 2】

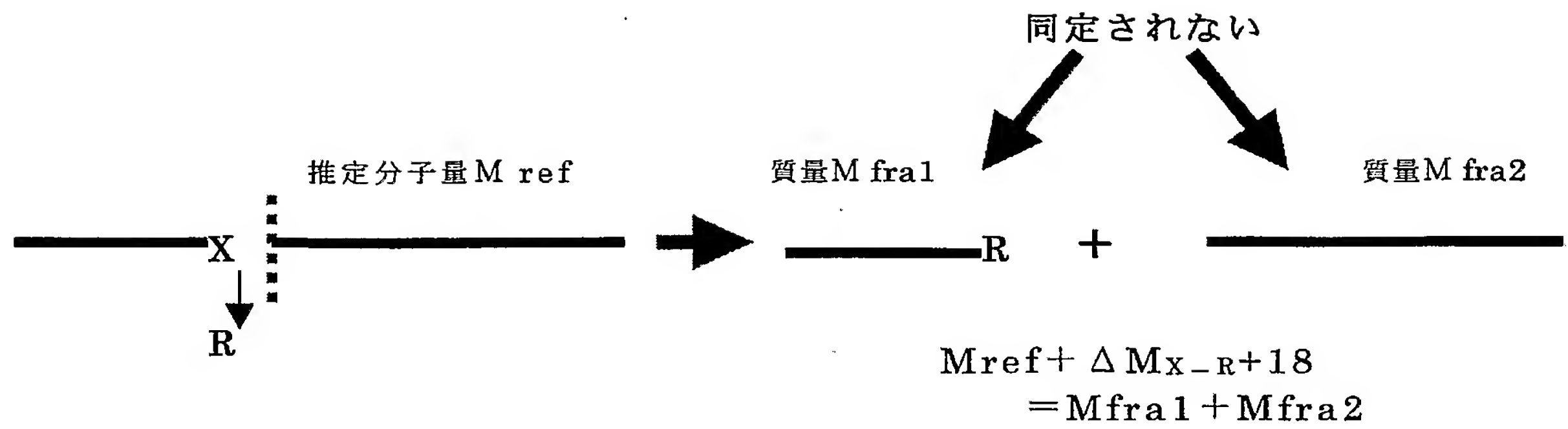


【図 3】

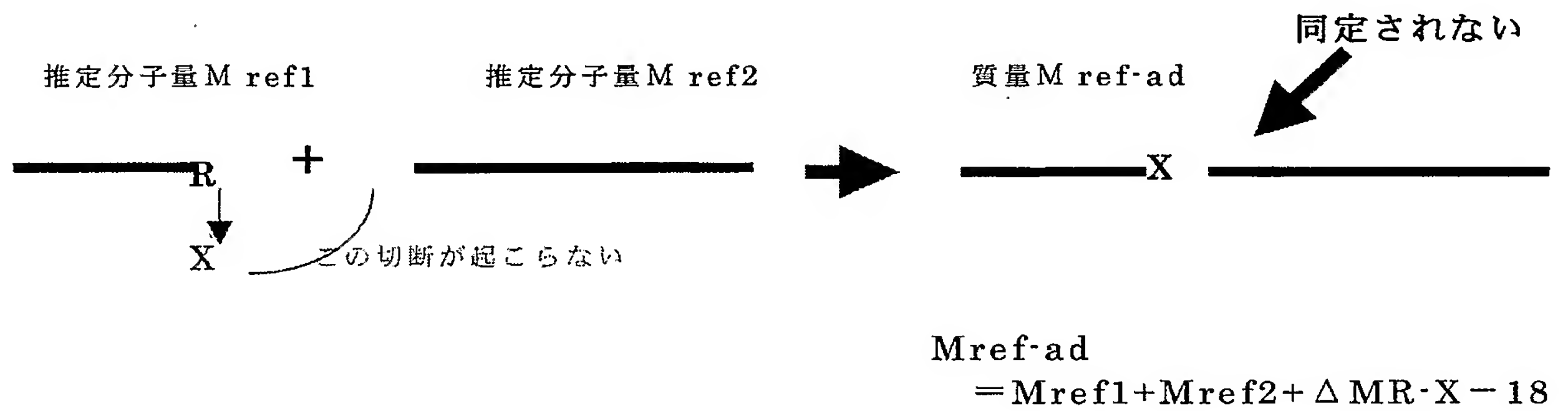
C 末切除型タンパク質



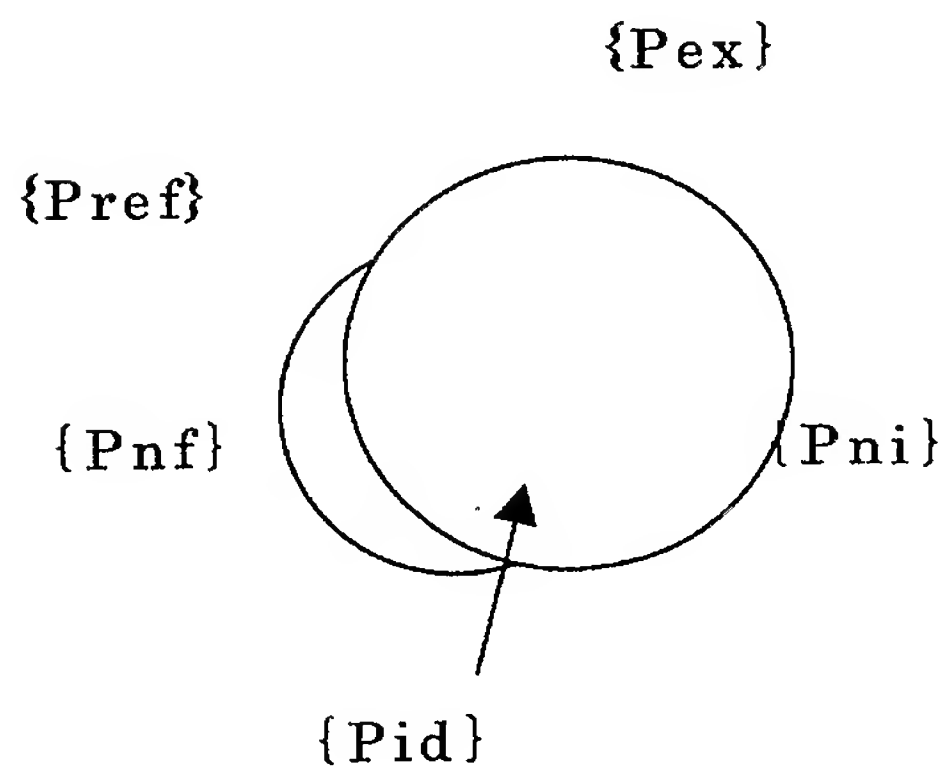
【図 4】



【図 5】



【図 6】





【書類名】 要約書

【要約】

【課題】 データ・ベース上の公知タンパク質をコードする遺伝子の塩基配列、ならびに、推断される完全長アミノ酸配列を参照し、解析対象タンパク質を部位特異的酵素消化したペプチド断片複数の質量分析結果に基づき、該対象タンパク質と同じゲノム遺伝子に由来する公知タンパク質、ないしは公知タンパク質変異体を高い確度で特定する手法の提供。

【解決手段】 部位特異的なタンパク質分解処理によるペプチド断片化を利用した、該対象タンパク質由来のペプチド断片分子量実測値と、公知タンパク質の推断完全長アミノ酸配列に基づき、推定されるペプチド断片分子量推定値とを対比し、一致する断片数の多寡、ならびに、公知タンパク質の一致断片アミノ酸配列の連続性、更には、不一致断片における変異推定を含む対比により、同定候補となる公知タンパク質を高い確度で特定する。

【選択図】 なし

特願 2 0 0 3 - 4 2 2 7 8 1

出 願 人 履 歴 情 報

識別番号

[ 0 0 0 0 0 4 2 3 7 ]

1. 変更年月日

1 9 9 0 年 8 月 2 9 日

[変更理由]

新規登録

住 所

東京都港区芝五丁目 7 番 1 号

氏 名

日本電気株式会社